













Volume 83 Issue 1 ISSN 0161-956X 2008

# *Peabody Journal of Education*

**CURRENT YR/VOL**

**Marygrove College Library  
8425 West McNichols Road  
Detroit, MI 48221**

Published in cooperation with the Peabody College of Vanderbilt University

ROUTLEDGE



**Routledge**

Taylor & Francis Group



# PEABODY JOURNAL OF EDUCATION

Camilla P. Benbow, Dean, Peabody College

James W. Guthrie, Editor

Michele Thompson, Coordinating Editor

Deborah M. Enright, Associate Editor

## EDITORIAL ADVISORY BOARD

Darlyne Bailey, *Teachers College, Columbia University*

R. Dale Ballou, *Peabody College of Vanderbilt University*

Sir Clive Booth, *Oxford Brookes University, England*

Mary M. Brabeck, *New York University*

Mark Bray, *The University of Hong Kong, PRC*

Louis A. Castenell, Jr., *The University of Georgia*

David Y. Chen, *Logistical Engineering University, PRC*

Lora Cohen-Vogel, *Florida State University*

Joseph P. Du Cette, *Temple University*

Victor Firsov, *Russian Academy of Teachers Retraining, Russia*

Juan Froemel, *LLECE/UNESCO, Chile*

Louis M. Gomez, *Northwestern University*

Gerardo M. Gonzalez, *Indiana University*

David W. Gordon, *Elk Grove Unified School District, California*

Eric A. Hanushek, *Hoover Institution, Stanford University*

Stephen P. Heyneman, *Peabody College of Vanderbilt University*

Phillip W. Jones, *The University of Sydney, Australia*

Kazuo Kuroda, *Waseda University, Japan*

James H. Lytle, *Trenton Public Schools, New Jersey*

Michael K. McLendon, *Peabody College of Vanderbilt University*

Ann L. Marcus, *New York University*

Bill McKinney, *Region IV Education Service Center, Texas*

Francois Orivel, *University of Bourgogne, France*

Penelope L. Peterson, *Northwestern University*

Michael J. Podgursky, *University of Missouri-Columbia*

Robert Slavin, *Johns Hopkins University*

Jandhyala B. G. Tilak, *National Institute of Education Planning and Administration, India*

---

**Abstracted and/or Indexed in:** PsycINFO/Psychological Abstracts; Linguistics and Language Behavior Abstracts; Contents Pages in Education; Sociological Abstracts; Education Index; Education Abstracts; Social Services Abstracts; EBSCOhost Products; Cabell's Directories; Family Index Database.

*Peabody Journal of Education* (ISSN: 0161-956X) is published quarterly in February, May, August, and November for a total of 4 issues per year by Taylor and Francis Group, LLC, 325 Chestnut Street, Suite 800, Philadelphia, PA 19106.

**US Postmaster:** Please send address changes to *Peabody Journal of Education*, Taylor & Francis Group, LLC, 325 Chestnut Street, Suite 800, Philadelphia, PA 19106.

### Annual Subscription, Volume 83, 2008

Print ISSN - 0161-956X, Online ISSN - 1532-7930

Institutional subscribers: US \$462, UK £277, €370

Personal subscribers: US \$58, UK £35, €46

Institutional and individual subscriptions include access to the online version of the journal. Institutional subscriptions include access for any number of concurrent users across a local area network. Individual subscriptions are single username/password only.

**Production and Advertising Office:** 325 Chestnut Street, Suite 800, Philadelphia, PA 19106. Tel: 215-625-8900, Fax: 215-625-8563. Production Editor: Michelle McCarthy

### Subscription offices

**USA/North America:** Taylor and Francis Group, LLC, 325 Chestnut Street, Suite 800, Philadelphia, PA 19106. Tel: 215-625-8900, Fax: 215-625-2940.

**UK/Europe:** Taylor and Francis Customer Services, Sheepen Place, Colchester, Essex CO3 3LP, UK. Tel: +44 (0)-20-7017-5544, Fax: +44 (0)-20-7017-5198.

For a complete guide to Taylor and Francis Group, LLC's journal and book publishing programs, visit our website: [www.taylorandfrancis.com](http://www.taylorandfrancis.com)

**Copyright © 2008 Taylor and Francis Group, LLC.** All rights reserved. No part of this publication may be reproduced, stored, transmitted, or disseminated in any form or by any means without prior written permission from Taylor and Francis Group, LLC. Taylor and Francis Group, LLC grants authorization for individuals to photocopy copyright material for private research use. The publisher assumes no responsibility for any statements of fact or opinion expressed in the published papers. The appearance of advertising in this journal does not constitute an endorsement or approval by the publisher, the editor, or the editorial board of the quality or value of the product advertised or of the claims made for it by its manufacturer.

**Permissions.** For further information, please visit: <http://www.tandf.co.uk/journals/permissions.html>

February 2008



## Introduction to the Special International Issue

Stephen P. Heyneman

*Department of Leadership, Policy, and Organizations  
Peabody College of Vanderbilt University*

It is axiomatic to suggest that the world of education is influenced by global trends. What may not be so obvious is the fact that some trends have deep historical roots, that progress can be stimulated by looking to see how other nations handle particular problems, and that global trends can have a positive effect. The articles in this issue illustrate each of these points.

John Smyth's article "The Origins of the International Standard Classification of Education" is an illustration of an important global effort to acquire a standard definition of educational institutions. Few of us may realize that statistics such as enrollment rates, the percentage of a population in higher education, or the percentage of students studying in vocational schools require a common metric. That metric is referred to as ISCED. This article recalls the earliest attempts to compile international educational statistics going back into the 19th century and retraces the steps that led up to the formulation and adoption of ISCED. In large part, it is a story of how international educational statistics came to be developed.

Much attention has been devoted to the problems of achieving universal basic education. In many low-income countries school tuition and other fees have been a significant barrier. Some scholars have asked why countries don't make primary education free. In their article "Implementing Free Primary Education Policy in Malawi and Ghana: Equity and Efficiency Analysis" Kazuma Inoue and Moses Oketch tell the story of two nations that eliminated tuition. They find that when the new policy is a political objective without attention to the required resources in teacher training and equipment, the result can increase both inefficiency and inequality. These two cases illustrate this point.

The principles that govern higher education admissions have shifted all around the world. In the article "Globalization and Implementation of an Equity Norm in Higher Education: Admission Processes and Funding Framework Under Scrutiny," Gaële Goastellec reports that these principles have shifted from "inherited" merit to a principle of equality of opportunity. Her article describes this shift and how the changes in policy are implemented. Goastellec examines how higher education "traduces" or "transcodes" the principle into practice when designing admission and funding policies. She concentrates on one of the consequences of the globalization of higher education, namely, the affirmation of equity as the key ingredient by which the organization and management of higher education systems will be judged.

The 15 new republics that emerged from the implosion of the former Soviet Union have many common educational characteristics, including a long history of centralized education finance. But all 15 nations have regional authorities and long traditions of local pride. The question is whether there will be any role for regional authority in the governance and/or financing of higher education. Rita Kasa, in her article titled "Aspects of Fiscal Federalism in Higher Education Cost Sharing in Latvia," finds that the answer is yes. Local authorities in Latvia have identified a role for themselves in higher education finance by helping to guarantee the loans of higher education students from their particular districts. However, the fact that regional authorities have identified a role does not mean that the role they have identified is strategically well formulated or consistent with national higher education objectives.

Globally, three demands characterize higher education: the demand for higher quality, the demand for higher access, and the demand for higher equity. Wherever public resources are limited, such as in East Africa, no nation has been able to meet these demands on the basis of public expenditures alone. Instead countries have had to seek financing from nonpublic sources, including tuition. But how can nations maintain their sense of equity in the face of rising tuitions? Many countries have responded to this dilemma by instituting "dual track" policies in which the most capable applicants are financed from public resources. The article by Marcucci, Johnstone, and Ngolovoi titled "Higher Education Cost-Sharing, Dual-Track Tuition Fees and Higher Education Access: The East African Experience" describes dual track policies in Tanzania, Kenya, and Uganda. The authors find what may have been anticipated given similar policies in Europe and North America, namely, that rewarding academic merit may not increase educational opportunity for the poor.

The quality of higher education is associated with economic development. In general, higher education quality is lower in countries with lower incomes/capita. What has been the effect of electronic technologies? Have the new technologies exacerbated or ameliorated these inequalities? In his article titled "Do Electronic Technologies Increase or Narrow Differences in Higher Education Quality



Between Low- and High-Income Countries?" Norman Clark Capshaw addresses this question. He constructs different answers to the question at the national level, the level of the individual institution, and the level of the classroom within specific institutions. He finds that usage of the Internet, and other computer technologies, in low- to middle-income countries is less than in high-income countries. But when specific enabling policies are put into place, the use of electronic technologies has the potential of ameliorating many of the international differences in higher education quality.

The role of education in fostering economic growth and social development is universally recognized. Although history places the provision of education firmly within national control, countries increasingly search outside national borders for alternative distribution frameworks. The World Trade Organization recently included education as service trade sector in the General Agreement for Trade in Services negotiations. Such activity increases debate about control as countries struggle to create policies that balance nationalism with economic responsiveness. The article titled "Compulsion, Craft, or Commodity? Education Services Trade in the Larger Context" by Brandyn Payne employs multivariate analysis to ask several questions. The first is whether trade openness in 162 countries was associated with openness to trade in education. The second is whether countries' commitments to lower barriers to education trade paralleled the strength of their commitments to lower barriers to all trade. She finds that countries with World Trade Organization education trade commitments have higher levels of general trade openness than those without education commitments. In lower middle income countries, education trade openness and general trade openness were positively related. When controlling for education, population, geography, and income, lower levels of education trade barriers were the single best predictor of countries' having made education commitments under General Agreement for Trade in Services. The question of whether international trade is 'good' for education is addressed in some detail. Her lesson might suggest that the debates over whether treating education as a tradable commodity is "bad" or "good" have failed to influence the authorities responsible for trade. If a nation is inclined to open itself to international trade, it will also do so in the field of education.

Special education is an increasing concern to educators. The proportion of children diagnosed with learning disabilities is on the increase, as are the resources needed for special education interventions. But how universal are these trends? Is the incidence of special education identical across societies? Are the interventions, judged to be necessary in one nation, considered equally necessary in another? In her article titled "Diagnosis, Treatment, and Educational Implications for Students with ADHD in the United States, Australia, and the United Kingdom," Sarah Schlachter responds to these questions. She finds that there are two definitions of ADHD used internationally and that the incidences, resources, and interventions used to address the problem differ dramatically from one environment to another.

These articles are representative of the conclusions from international education policy more generally. The Smyth article teaches us that little progress can be made without a consensus on common structures necessary for comparison. The Schlachter article reminds us that a nation which designs policy by solely utilizing its own experience may risk creating unnecessary distortions. The Kasa article; the Inoue and Oketch article; and the Marcucci, Johnstone, and Ngolovoi article suggest that, however compelling, international reform norms can be problematic unless attention is paid to the local implementation requirements. The Payne and Capshaw articles suggest that there is progress in international education and that nations with open policies intelligently administered may well be future leaders of us all.



# The Origins of the International Standard Classification of Education

John A. Smyth  
(Formerly UNESCO, Paris, France)<sup>1</sup>

This article recalls the earliest attempts to compile international educational statistics going back into the 19th century and retraces the steps that led up to the formulation and adoption of the International Standard Classification of Education in both its original and revised versions. It is in large part the story of how international educational statistics came to be developed. Source documents that have long been out of print and/or not easily accessible to readers outside UNESCO are quoted at length.

Differences between countries in the organization and contents of their education are reflected in different national terminologies and classifications of education, thus making it difficult to compile internationally comparable educational statistics. The problem was described many years ago by distinguished comparative educationist Nicholas Hans. Hans (1933) also suggested a possible solution:

In comparing educational systems the first difficulty is that of classification and terminology. The same terms used in different countries often denote quite different institutions. Thus, “*école secondaire*” in France, “*sekundarschule*” in Switzerland and “secondary school” in England are not really synonymous terms. The French “*collège*”, the English “college” and the American “college” are different institutions with divergent standards. The same applies to the term “middle school” and to the terms “gymnasium” and “lyceum”. The German and the Dutch “lyceum”, for instance, do not mean the same thing at all. *The problem can be solved only by using*

---

<sup>1</sup>John Smyth is a former United Nations Educational, Scientific and Cultural Organization (UNESCO) official (1972–2000). The views expressed in the article are the author’s and do not represent those of UNESCO.

The assistance of UNESCO’s Archives Service and the staff of the Documentation Centre of the International Bureau of Education, Geneva, Switzerland, in locating relevant reference documents, is gratefully acknowledged.

Correspondence should be sent to John A. Smyth, 7 rue Pierre Villey, Paris 75007, France. E-mail: [jsmyth@club-internet.fr](mailto:jsmyth@club-internet.fr)

*an artificial terminology which can be applied uniformly to all countries.* The second difficulty is to distribute schools in accordance with this accepted terminology. Every country has its own classification of schools, and knowledge of the curriculum and the standards of each type is necessary in order to place them correctly. In the third place, statistics of the age of the pupils are often lacking and are nearly always insufficient. (p. lxxxviii)

Hans did not give any examples of an “artificial terminology” that could serve his purpose, but his proposal anticipated current practice. The first international classification of education to make use of an artificial terminology was developed by the United Nations Educational, Scientific and Cultural Organization (UNESCO) after the Second World War. It was contained in a formal “Recommendation concerning the International Standardization of Educational Statistics” adopted by the organization’s Member States in 1958 (UNESCO, 1958). Subsequently, UNESCO developed a classification known as the International Standard Classification of Education (ISCED), which came to be incorporated in a “Revised Recommendation” adopted by Member States in 1978 (UNESCO, 1976, 1978). A revised version of ISCED adopted in 1997 currently serves as the classification framework for the international educational statistics compiled by UNESCO and other international organizations such as the Organization for Economic Cooperation and Development (OECD) and the Statistical Office of the European Union (EUROSTAT; UNESCO, 1997).

This article recalls the earliest attempts to compile international educational statistics going back into the 19th century and retraces the steps that led up to the formulation and adoption of ISCED in both its original and revised versions, although it does not undertake to review the current (revised) version as such. It is in large part the story of how international educational statistics came to be developed. Source documents that have long been out of print and/or not easily accessible to readers outside UNESCO are quoted at length.

## DEVELOPMENTS BEFORE THE SECOND WORLD WAR

In most countries, the compilation of national educational statistics has traditionally been a task assumed by the public authorities responsible for education. Thus, the systematic collection and collation of international educational statistics could not easily be organized until there came into being an international body responsible for fostering intergovernmental cooperation in the exchange of information and experience in the field of education, including cooperation in compiling international educational statistics. This did not happen until the establishment of UNESCO after the Second World War. Before then, attempts to

compile international educational statistics had largely foundered on the question of international comparability.

As early as the 1850s, statisticians in a number of European countries had come to recognize education as a field of statistical inquiry that could benefit from the exchange of professional experience among statisticians of different countries, as was beginning to happen in other more established fields of inquiry such as population statistics. Indeed, education was one of 11 branches or fields of statistics separately identified for discussion at the first International Statistical Congress held in Brussels in 1853, and it was featured from time to time in the programs of subsequent congresses up until the First World War. The International Statistical Institute (ISI), which assumed the responsibility for organizing these congresses after it came into existence as a professional association of statisticians in 1885, took an active interest in this field, but the earliest studies were mostly single country studies, and such international educational statistics as were compiled by individual scholars and researchers were ad hoc compilations drawing on published data available for selected countries.<sup>2</sup> Before the First World War there did not exist an international organization or mechanism capable of compiling international educational statistics.

International interest in educational statistics emerged after the war as the result of efforts by certain countries to promote "international intellectual cooperation." During the peace negotiations that led up to the Treaty of Versailles, following the end of the war, various proposals were floated for the creation of international bodies that might help to ease tensions between countries and contribute in different fields toward the improvement of international understanding and the strengthening of peace. Agreement was reached on the establishment of the League of Nations as well as several other bodies, such as the International Court of Justice, and the International Labour Bureau, forerunner of the International Labour Office (ILO). In 1921 the assembly of the League of Nations adopted a proposal submitted by France calling for the establishment of an International Commission on Intellectual Cooperation to advise the league's council on measures that governments could take with a view to stimulating international intellectual cooperation in furtherance of the league's overall aims and objectives.<sup>3</sup> Hopes were expressed

---

<sup>2</sup>With the founding of ISI, the International Statistical Congresses were revived as biennial sessions of the General Assembly of ISI. Photocopies of early papers by ISI members concerning educational statistics are reproduced in a two-volume, limited circulation compendium prepared by ISI in 1995 (ISI, 1995). Particularly noteworthy for the scope of its coverage of the published data available for different countries, and for its acknowledgment of the problems of cross-national comparability, is an early paper (reproduced in the ISI compendium) on primary education statistics (by E. Levasseur, 1893).

<sup>3</sup>It was envisioned that the commission would meet for a month each year in Geneva and would be composed of a dozen eminent international figures in the sciences and humanities appointed by the



in the assembly and council that the commission would give consideration to matters such as international cooperation among scientific researchers; international relations between universities, particularly in respect of teacher and student exchanges and the mutual recognition of degrees and diplomas; the international circulation of scientific publications; intellectual property rights; the condition of libraries; and the development of international bibliographies in the sciences and humanities. In its decision to set up the commission, the assembly did not provide for the commission to have its own secretariat, which would have been tantamount to setting up another international agency like the International Labour Bureau, which the assembly wished to avoid.

To ensure that there would be some follow-up of the commission's recommendations, France decided in 1926 to set up in Paris the International Institute of Intellectual Cooperation (IIIC) as a kind of executive arm for the commission (IIIC, 1946). The commission was charged by the league with overseeing the institute's work, but although the institute was conceived by its statutes as an international organization "independent of the authorities of the country in which it is placed" (Article 3) and had several nationalities present on its board of management and staff, it did not have the status of an intergovernmental body or technical organization of the League of Nations, like the International Labour Bureau, and was in practice to carry out its activities mainly in cooperation with nongovernmental organizations. The league did not provide any funds for its operations, which were largely financed by France, although certain other countries as well as private bodies such as the Rockefeller Foundation and the Carnegie Endowment volunteered modest financial support.

Significantly, one of the institute's earliest initiatives was to set up a joint committee with the ISI at the end of 1926 to consider the needs for international statistics on "the principal manifestations of intellectual life in the different countries." In its report a year later the Joint IIIC-ISI Committee laid out a comprehensive set of model statistical tables for the presentation of national statistics on education, science, and culture that could at the same time serve as a common cross-national basis for the compilation of international statistics in these fields (March, 1928). A distinction was drawn between statistics that it would be desirable to collect annually, such as student enrollments, and those that could be collected at 5-year intervals because they changed little from year to year, such as the number of museums. Half the tables concerned education. The tables concerning science and culture covered various fields ranging from scientific research establishments to museums and archives, historic and artistic monuments, book production and publishing, theatres, concerts, cinema, radio broadcasting, patents and inventions, and employment in the liberal professions.

---

council to sit in a personal capacity and not as government representatives. On the origins and work of the International Commission on Intellectual Cooperation, see Renoliet (1999).

The education tables were classified under six headings: higher education (universities), secondary education, primary teacher education (normal schools), primary education, adult education, and specialized education (agricultural education, technical education, etc.). Although the Joint IIIC–ISI Committee was aware of the problem of international comparability, it did not put forward any rules or criteria for countries to follow in classifying their educational statistics under the various headings. This did not prevent ISI’s General Assembly from adopting a resolution calling for the preparation of national statistics in accordance with the model tables, and for the preparation, “in collaboration with ISI,” of a “trial” set of international statistics based on already-published data available for various countries, taking into account national differences in the relevant definitions, legislation, and administrative practices.<sup>4</sup>

The Joint IIIC–ISI Committee’s model statistical tables for education represent the first-ever attempt by an authoritative international body to conceive of a statistical classification of education that could be applied internationally. Whether the classification could be meaningfully applied was another matter, however, for it depended on whether the data assembled by any two countries for a given table—for example, figures of primary school enrollments—could be considered as comparable, that is to say, measures of the same thing in both countries. (How to interpret the figures if the duration of primary schooling is different in the two countries? Or if the content of primary schooling is different?)

There was no follow-up by either IIIC or ISI. IIIC in particular did not have the institutional mandate or in-house technical capacity that would have been needed if it were to carry out a program of compiling international statistics in its fields of interest, whereas ISI, although able to advise on such a program, was not a statistical agency that could assist in an operational sense in carrying it out. In any event, IIIC never attempted to put together a trial set of international statistics as recommended by the Joint Committee.

The Joint Committee’s report did not disappear completely from sight; it was recalled after the Second World War when a program of international statistics to be carried out by UNESCO was being considered. In the meantime, another body set up at around the same time as IIIC, the International Bureau of Education (IBE) in Geneva, started to compile international educational statistics.

IBE was originally founded in 1925 as an offshoot of the School of Educational Sciences of the University of Geneva with the goal of fostering the exchange of information and experience among educational researchers. In 1929 it adopted new statutes aiming to involve in its work the public authorities responsible for education in different countries, and specifically providing for “any government, public institution or international nongovernmental organization” to become a member of

---

<sup>4</sup>The assembly’s resolution is reproduced in March (1928, p. 638).



IBE on payment of an annual contribution to the institute's budget.<sup>5</sup> The eminent Swiss psychologist and educationist, Jean Piaget, was appointed Director. Under the new statutes, IBE became a 'centre of educational information' with a mandate 'to collect documentation on educational research and its applications, and ensure a wide-ranging exchange of such documentation and information in order that each country will feel stimulated to benefit from the experience of others'. The first members of IBE were the Republic and Canton of Geneva, the University of Geneva's School of Educational Sciences, the Governments of Ecuador, Egypt and Spain, and the Ministries of Public Education of Czechoslovakia and Poland.

The institute soon became very active in carrying out international surveys on selected educational topics by means of questionnaires addressed to the national educational authorities and leading institutions of educational research in different countries, whether or not they were actually members of IBE. The number of countries that were ready to participate in the institute's various surveys evidenced a strong latent demand among educational leaders and policymakers at that time for opportunities to exchange information and experience on the development of their education systems. The first surveys covered topics such as the practice of self-government in schools, the contribution of children's books to the spirit of international cooperation, and the teaching of child psychology in normal schools (teacher training colleges), the latter survey drawing responses from the national authorities and/or specialized institutions in 27 countries—the majority in Europe but also including non-European countries such as Argentina, Australia, Canada, Egypt, New Zealand, Palestine, South Africa, the United States, and Uruguay. In 1931–32, 53 countries responded to a survey on the organization of the public education system.

In 1931 the institute's Governing Council started the practice of inviting its national representatives to present reports on recent educational developments in their countries for discussion at the council's annual meeting. In 1932 the invitation was extended to any country that might be interested, whether or not it was a member of IBE. Thirty-five countries responded, and the institute put their reports together in a volume that was to become the first issue of the *International Yearbook of Education* (Bureau International d'Éducation, 1933). Encouraged by the response, the institute in the following year convened the first of what was to become a series of annual sessions of the International Conference on Public Education.

With the publication of its first *International Yearbook*, the institute found itself confronted with the question of how to handle the national educational statistics that were included in most of the country reports, in particular whether it should

---

<sup>5</sup>On the origins and early work of IBE, see Rossello (1943) and his "Historical Note" (Rossello, 1970).

attempt to compile international tables based on these statistics. The issue was raised by the representative of the University of Geneva's School of Educational Sciences, M. Dottrens, at the institute's Governing Council meeting of July 12–13, 1933 (Bureau International d'Éducation, 1933). Dottrens expressed the hope that the second (1934) edition of the *Yearbook* would include “summary” statistical tables—in effect, international tables—that would “recapitulate” the national statistics provided in the country reports, but Piaget was sceptical (“Experience has shown us that it is extremely difficult to establish comparable statistics”), and the idea was dropped, at least temporarily. Beginning with the 1937 edition, the *Yearbook* included international tables, as Dottrens had suggested.

Piaget was not the first educationist to be sceptical of the possibility of compiling internationally comparable educational statistics. Isaac Kandel (1925), the distinguished comparative educationist at Teachers College, Columbia University, New York, had expressed a similar scepticism a decade earlier in his introduction to the first edition of the *Educational Yearbook of the International Institute of Teachers College*, a publication which in many respects anticipated IBE's *Yearbook*:

One difficulty which will be readily recognized presented itself in the preparation of this volume, and that is in the field of statistics. Nomenclature, while it is not standardized in education, is readily comprehended from the context. In the case of international statistics uniformity of standards are not yet available. Frequently statistics are not available for the same year; in each case the latest published by the official departments have been used. In the case of financial statistics the difficulty has been aggravated by the fluctuating values of the post-War period. No attempt had been made to reduce such figures to a common standard. In course of time this will become increasingly possible. Before that time is reached, however, the realization of these difficulties may, it is hoped, lead to the development either of some international agency for the collection and interpretation of such statistics, or for the general acceptance of some system of educational records and reports that approach a semblance of uniformity. (p. x)

The Teachers College *Educational Yearbook* therefore did not present any international statistical tables. Kandel's hope that “some international agency” would one day emerge to take responsibility “for the collection and interpretation of [educational] statistics” would be realized after the Second World War in the form of UNESCO.

Piaget's and Kandel's scepticism of the possibility of compiling internationally comparable educational statistics was not shared, at least to the same degree, by the editors of another *Yearbook* series, *The Year Book of Education*, which was launched by a group of leading educationists in the United Kingdom in 1932, a year before IBE's *International Yearbook* series. Like the Teachers College *Educational Yearbook*, *The Year Book of Education* contained individual country chapters on recent educational developments prepared by educationists in the countries



concerned. In addition it presented a set of international tables of student enrollments, numbers of teachers and educational expenditures, and a chapter entitled "Comparative Statistics" prepared by the comparative educationist Nicholas Hans (1933), setting out the approach adopted in compiling the tables (pp. lxxxviii–xc).<sup>6</sup> Hans's general observations on the problem of international comparability were quoted at the beginning of this article. In compiling the tables, however, his choices of both terminology and the unit of classification (the educational institution) did not wholly free the tables from ambiguities of interpretation:

In these tables, therefore, the classification adopted is based neither on official terminology nor on age, but on the functions of each type of school. Thus *pre-school institutions* include Kindergartens, Nursery Schools and other independent institutions which do not form an integral part of the primary school. The Infant departments of primary and secondary schools are not included. The ages of pupils in pre-school institutions vary in accordance with legislation on compulsory education. In the majority of countries the ages are from 3 to 6 or from 3 to 7, but in Russia and Latvia they are from 3 to 8. The *Primary School* does not only include the Infant, Junior and Senior Departments, but also the advanced divisions which are not separated into independent intermediate schools. The ages again vary. In the majority of countries they are 6–14 or 7–14, but in Russia, for instance, they comprise only the ages 8–12, in Latvia 8–14, in France and Holland 6–13 and in Hungary and Japan 6–12.

*Intermediate Schools*, though catering to different ages in different countries, have the same function of giving a post-primary education of a non-vocational character (i.e. an education not specialized according to vocations), to children not proceeding to universities. The preparatory departments of such schools are included if they form an integral part of them. Thus, for instance, the Austrian intermediate school (*Hauptschule*) comprises only the ages 10–14, whereas the Russian seven-year school extends from 8–15.

*Secondary Schools* include schools definitely preparing for universities and higher technical colleges. In the United States and the British Dominions they are separate institutions without preparatory or intermediate departments; in other countries they combine intermediate and secondary education, and in some countries they include primary grades as well. On the other hand, where, as in Canada, education is classified

---

<sup>6</sup>The international tables covered 24 countries: Australia, Austria, Belgium, Canada, Czechoslovakia, Denmark, England, France, Germany, Holland, India, Irish Free State, Italy, Japan, New Zealand, Norway, Poland, Scotland, South Africa, Soviet Russia, Spain, Sweden, Switzerland, and the United States. Starting with the 1937 edition, *The Year Book* was published by Evans Brothers "in association with the University of London Institute of Education." After a hiatus in 1941–1947 due to the war, *The Year Book* resumed publication in 1948 but without including tables of international educational statistics. Teachers College stopped publishing its *Educational Yearbook* after the 1952 edition and in the following year entered into an arrangement with the University of London Institute of Education for jointly preparing the Evans Brothers *Year Book*.

by grades rather than by schools, we have endeavoured to include pupils of secondary grade in primary schools in the Secondary column. As a rule secondary schools include ages up to 18, but in Norway, for instance, up to 19.

*Vocational Schools* include all post-primary institutions and day and evening classes preparing for a definite vocation. Teachers' training colleges, if not of university standard, are included here. Adult schools and classes are excluded.

*Higher Education* includes universities and other higher institutions of academic standard, whose curricula are based on complete secondary education. In some countries, however, the United States, for instance, they include colleges of questionable academic standard. It is evident that this artificial classification is not adequate, but it may serve until some international authority undertakes the task of revision. (lxxxviii–ix)

Like Kandel earlier, Hans foresaw that “some international authority” would need to take responsibility for developing a classification less prone to ambiguous interpretation. His idea of a classification based on “the functions of each type of school” partly anticipated ISCED, except that ISCED was to utilize the “educational programme” rather than the school or educational institution as the unit of classification. In choosing the school as the unit of classification, Hans retained the ambiguities of interpretation inherent in applying regular English terms for different kinds of schools cross-nationally. Although he was aware of the need for “an artificial terminology which can be applied uniformly to all countries” (as quoted at the beginning of this article), he did not actually devise such a terminology.

*The Year Book of Education's* initiative in presenting international tables may well have persuaded IBE to go back on its earlier decision and to include such tables in the 1937 edition of its *International Yearbook*. Seven international tables were included:

- Budgets of Ministries of Public Education
- Number of primary schools
- Number of primary school pupils
- Number of primary school teachers
- Number of secondary schools
- Number of secondary school pupils
- Number of secondary school teachers

For each table, data were shown separately for each of the five school years 1932/33 to 1936/37. The figures in the Budgets table were expressed in national currency units; no attempt was made to convert them into a common unit. In the other tables, figures for public and private schools were shown separately.

Because the *International Yearbook* did not explicitly claim that the figures for the different countries in each table were comparable, the onus of interpreting the



tables basically fell on the reader. Unlike *The Year Book of Education*, IBE's *International Yearbook* did not present any explanatory tables showing, for example, the normal durations of the various kinds of education in the different countries or the age ranges of compulsory education. Readers of IBE's *International Yearbook* who were already familiar with the education provided in one or more countries other than their own would to some extent have been able to assess the degree to which the figures for their own country in a given table could be meaningfully compared with the corresponding figures for other countries. Few if any readers, however, would have been able to do this in respect of all the countries in the table. The utilization of a standard terminology ("primary school," "secondary school") implied that the figures for the different countries in any table concerned similar institutions, but in the absence of explanatory notes the nature of the similarity could only be speculative (curricula? duration? age of pupils at entry?). Thus, the *International Yearbook's* tables were even more open to ambiguous interpretation than the tables compiled by Hans for *The Year Book of Education*.

Up until the Second World War, therefore, the question of international comparability in educational statistics had not been resolved. There had been, however, a considerable expansion of international cooperation in the exchange of information and experience among leading educationists of different countries, and a growing understanding of the nature of the difficulties involved in compiling international educational statistics.

### UNESCO'S EARLY WORK ON INTERNATIONAL EDUCATIONAL STATISTICS 1946–1958<sup>7</sup>

International opinion at the end of the Second World War was more favorably disposed toward intergovernmental cooperation in the field of education than it had been at the end of the First World War. Thus, UNESCO came into existence in 1946 with a broad mandate to promote international cooperation in education as well as in science and culture.<sup>8</sup> At the same time, IIIC, which had been briefly revived toward the end of the war, was closed down; many of its activities were taken over by UNESCO. IBE, on the other hand continued with much the same program of activities that it had before the war, while reaching agreement with UNESCO on

---

<sup>7</sup>Highlights of UNESCO's early work on international educational statistics are recounted in UNESCO publications (UNESCO, 1955, pp. 47–56; UNESCO, 1961a, pp. 26–35) and in Kappel (1966, pp. 661–668).

<sup>8</sup>UNESCO was established "for the purpose of advancing, through the educational and scientific and cultural relations of the peoples of the world, the objectives of international peace and of the common welfare of mankind for which the United Nations Organization was established and which its Charter proclaims" (preamble of UNESCO's constitution). On the establishment of UNESCO, see Valderama (1995).

jointly reconvening the International Conference on Public Education under a new title, the International Conference on Education.

Article VIII of UNESCO's constitution specifically provided for cooperation with Member States in respect of "statistics relating to their educational, scientific and cultural activities." Thus, UNESCO assumed the role of lead organization within the United Nations system responsible for international statistics in these fields. A statistical office was established in April 1950.

In its early work on educational statistics the new office's main concerns were twofold: to build up a database of the national statistics then available while setting in motion a process of expert consultation on the question of international comparability and standardization. The work on the database dovetailed with a larger programme of UNESCO's Education Department aiming to establish a profile of the worldwide provision of formal education and the incidence of illiteracy. Initial results of this program were the publication of a *World Handbook of Educational Organization and Statistics* (UNESCO, 1952) containing a core of statistics, descriptive text, diagrams, glossaries, and bibliographies on the education systems of 57 countries, and a monograph, *Progress of Literacy in Various Countries* (UNESCO, 1953), which brought together in one place for the first time historical data concerning literacy as reported in the national censuses of 23 countries going back to 1900.<sup>9</sup>

In these early publications statistical data were presented essentially in the countries' own terms, as at that time there were no internationally agreed definitions and rules for the classification of education. Although most countries distinguished between primary, secondary, and higher education, the interpretation of these categories varied from country to country. In the *Progress of Literacy* monograph, attention was drawn to the various definitions of literacy utilized by countries in their national censuses (e.g. "can read and write," "can read only," "can read and write Spanish," and so on). The need for international standard definitions, whether of literacy or different categories of education, was felt particularly when plans were made for the publication of a periodic World Survey of Education. The first edition of the World Survey, published in 1955 (UNESCO, 1955), was largely an expanded version of the 1952 *Handbook* aiming to cover virtually all the world's countries and territories,<sup>10</sup> but its purposes were both synthetic and normative:

The Universal Declaration of Human Rights, adopted unanimously by some 50 nations at the third session of the UN General Assembly on 10 December 1948, states in Article 26 (1): "Everyone has the right to education. Education shall be free,

---

<sup>9</sup>For a critical review of UNESCO's work on international literacy statistics during the period 1950 to 2000, see Smyth (2006).

<sup>10</sup>The word *territories* was utilized at that time to refer to countries that had not yet gained independence.

at least in the elementary and fundamental stages. Elementary education shall be compulsory. Technical and professional education shall be made generally available and higher education shall be equally accessible to all on the basis of merit." This is the educational profession of faith of the world today. But for a full understanding of the goals humanity has set itself one needs to place beside the Universal Declaration a "situation report" on the present state of educational affairs—the purpose being constructive, to reveal the size of the task ahead, and not simply to reflect negatively on how far reality falls short of the ideal. This chapter is intended as such a survey; or rather, considering the imperfect information now available, it may serve as the first outline for such a survey. In due course the gaps will be filled in, the techniques improved, and a report may later emerge showing, more fully, both facts and trends: where education is, how it is moving. (UNESCO, 1955, p. 13)

In adopting Article 26 of the Universal Declaration of Human Rights as a standard on which to assess "the present state of educational affairs" in the world, the World Survey in effect acknowledged a compelling normative basis for assembling comparable educational data from different countries. For the educational statistician the challenge was not merely the existence of gaps in statistical information for so many countries at that time; it was also the lack of international comparability in the statistics that were available. The World Survey attempted to make a broad, tentative assessment, estimating that around half of the world's adult population 15 years of age and older was illiterate and had therefore in effect been denied the right to education (pp. 13–16) and that "at least half of the world's children [between 5 and 14 years of age] were not receiving any kind of school education" (p. 17). This was the first time that any official body had presented such estimates.

The second of the two main thrusts of the new statistical office's work—the setting in motion of a process of expert consultation on the standardization of national educational statistics—was well under way at the time when the World Survey was put together:

An Expert Committee on Standardization of Educational Statistics met in November 1951, under the chairmanship of Professor P. J. Idenburg (Netherlands), and proposed a minimum set of definitions, classifications and tabulations of statistics on illiteracy and education. The report of this committee, and a working paper by the Secretariat, were sent to all Member States for comments. The subject was also presented at the twenty-eighth session of the International Statistical Institute (Rome, 1953), the eighth session of the United Nations Statistical Commission (Geneva, 1954), and the third Inter-American Statistical Conference (Petropolis, Brazil, 1955).<sup>11</sup> (UNESCO, 1961a, p. 33)

---

<sup>11</sup>Professor Idenburg was well known internationally for his interest in educational statistics; several of his papers are reproduced in the ISI Compendium cited in footnote 2. The paper presented by UNESCO at the 28th Session of ISI held in Rome in 1953 contains a resume of the Expert Committee's



The extensive and lengthy process of consultation on the Expert Committee's proposals was necessitated both by the number of interested parties and by the novelty of the proposals themselves. For the first time, countries were presented with a set of definitions and rules for classifying education designed to facilitate the international comparability of their educational statistics. Although the Joint IIIC–ISI Committee before the war, as was noted earlier, had proposed a set of standard tables as well as a classification of education (higher education, secondary education, etc.), it had not specified the definitions and rules (criteria) of classification that countries should follow if their national tabulations were to be internationally comparable.

The definitions proposed by the Expert Committee were as follows:

The Committee recommends the following definitions to be applied to statistics on education:

- (i) *Compulsory school age population* is the population between the age limits of compulsory full-time education, apart from exceptions as provided in the law of each country (State, province, etc.).
- (ii) In countries where education is not compulsory, *the school age population* includes all children within the usual ages of entering and completing the typical primary school according to the practice of that country.
- (iii) *A government financed school* is one which is basically financed from official (federal, State or local government) sources, whether or not supplemented by fees or incidental gifts.
- (iv) *A government aided school* is one which is partly financed from official sources.
- (v) *An Independent school* is one which receives no financial support from official sources.
- (vi) *A School* is a group of pupils or students organized as a single educational unit under one or more teachers with an immediate head.
- (vii) *A Class* is a group of pupils who are usually instructed together by a teacher—not necessarily the same teacher all the time.
- (viii) *A Grade* (standard, form, etc.) is a stage on the educational ladder of one school year's (or academic year's) duration.
- (ix) *A Student* or pupil is a person enrolled for full-time or part-time education at any level.
- (x) *A Teacher* is a person directly engaged in educating a group of pupils or students.

(Note: The number of teachers at any level of education below higher education is the number of full-time teachers, i.e. teachers engaged during the normal school day

as provided in the timetable of that school plus the full-time equivalent of part-time teachers.) (UNESCO, 1954, pp. 513–514)

Because a term needed to be defined only if it were to be utilized in a table, the majority of the committee's definitions were not strictly necessary. The international tables recommended by the committee concerned only four terms: *school*, *class*, *student*, and *teacher*.<sup>12</sup> However, it is likely that the committee anticipated that the other terms would come into use at a future date.

The classification proposed by the committee distinguished between four broad categories of education. The first basically encompassed formal school and university education, broken down into four "levels," whereas the fourth was essentially a residual category:

The Committee recommends that for purposes of international reporting schools should be classified as far as possible by level and type as follows:

- (a) Education, by level:
  - (i) *A school of the first level* (e.g. nursery school, kindergarten, infant school) provides education for children who are not yet ready to enter a school of the second level.
  - (ii) *A school of the second level* (e.g. elementary school, primary school) provides basic instruction in tools of learning, as well as education for the social and emotional development of the children.
  - (iii) *A school of the third level* (e.g. middle school, secondary school, high school) provides general or specialized instruction more advanced than that given at the second level. As to schools of the third level the education is subdivided into: (a) general education, which does not aim to prepare the pupils for a certain profession or trade; (b) vocational education, which aims to prepare the pupils directly for a certain profession or trade.
  - (iv) *An institution of the fourth level* is one which requires, as a minimum of admission, a certificate of completion of a school of the third level or its equivalent (e.g. an entrance examination). Institutions of this level include universities and higher professional schools.
- (b) Teacher education
- (c) Special education is all general or vocational education given to physically or mentally handicapped, socially maladjusted, retarded or backward persons.
- (d) Supplementary education includes all education not included elsewhere (e.g. adult education). (UNESCO, 1954, pp. 514–515)

---

<sup>12</sup>The tabulations are listed at UNESCO (1954, p. 515). The committee considered them as representing "a minimum programme . . . of tabulations of educational statistics for international purposes" (p. 515).

This was the first time that an authoritative group of international experts had proposed that education and/or schools and educational institutions could—indeed should—be classified for statistical purposes by level. The committee did not define the term *level* as such but evidently had in mind successive stages of a process of instruction running from infant–nursery school up to university. It nominated the primary–elementary school as “a school of the second level” and then defined “a school of the first level” as one that “provides education for children who are not yet ready to enter a school of the second level.” “A school of the third level” was then defined as one that “provides general or specialized instruction more advanced than that given at the second level,” whereas “an institution of the fourth level” was simply one that “requires as a minimum of admission, a certificate of completion of a school of the third level or its equivalent.” Thus, anchored on “a school of the second level,” all the “levels” were linked together like the steps on a ladder. This was certainly “an artificial terminology which can be applied uniformly to all countries”—Nicholas Hans’s proposed solution of the cross-national comparability problem—but the solution is valid only to the extent that different countries’ “second-level” schools, on which the hierarchy of levels in each country is anchored, can be considered as comparable.

Not all of the Expert Committee’s proposals met with the approval of the various bodies consulted by UNESCO. Several problems with the recommendations were highlighted by the organization’s statistical service in its presentation at ISI’s 28th session in Rome in 1953 (UNESCO, 1954, pp. 515–516). One problem was the committee’s nomenclature for the levels, although the levels as such were broadly accepted: schools falling between the primary–elementary level and the level of higher education were to be classified as the third level, whereas in many countries they would be considered “secondary” schools. Another problem was the proposed classification of teacher education as distinct from other types of formal education: In many countries it would be difficult in practice to separate teacher-training institutions from other types of educational institutions, especially at the postsecondary level. Also questionable were the committee’s definitions of “compulsory school-age population” and “school-age population,” which “would seem to be quite inadequate for international comparability, unless some arbitrary chronological age-group, such as 5–14 years inclusive, were adopted as a common denominator” (UNESCO, 1954, p. 516). Problematic too were the committee’s definitions of “government-financed” and “government-aided” schools, because it was unclear how countries could in practice distinguish between schools that are “basically financed from official sources” and “partly financed from official sources.” Moreover, the committee’s definition of “an independent school” did not cover those schools and educational institutions which in some countries were legally and administratively independent even though they received financial aid from official sources.



Evidently most of these problems could be ironed out by relatively straightforward revisions of the Expert Committee's proposals. There was a broad consensus among countries that international agreement on a revised set of proposals would be feasible. Accordingly, in 1956 UNESCO's General Conference authorized the Director-General to convene an intergovernmental committee of experts with a mandate to prepare a draft international agreement for submission to the conference at its 1958 session. The result was the adoption by the conference of the "Recommendation concerning the international standardization of educational statistics" (UNESCO, 1958).

### THE 1958 RECOMMENDATION

The 1958 Recommendation marked the first time that States agreed on a common set of definitions and principles of classification specifically designed for international reporting of their educational statistics.<sup>13</sup> In drawing up the Recommendation, the intergovernmental committee considered a broader range of educational statistics than had been considered by the 1951 Expert Committee, which had focused mainly on statistics of educational institutions. The recommendation's provisions were set out under four headings:

- I. Statistics of Illiteracy
- II. Statistics of the Educational Attainment of the Population
- III. Statistics of Educational Institutions
- IV. Statistics of Educational Finance

Under Statistics of Illiteracy,<sup>14</sup> the 1958 Recommendation presented two definitions:

- (a) A person is *literate* who can with understanding both read and write a short simple statement on his/her everyday life.
- (b) A person is *illiterate* who cannot with understanding both read and write a short simple statement on his/her everyday life (UNESCO, 1958, para. 1).

Many countries were already using definitions similar to these in their national censuses, following an earlier recommendation of the United Nations Population

<sup>13</sup>See the preamble of the General Conference resolution in the appendix.

<sup>14</sup>During most of the period of 1950 to 2000, UNESCO's publications of international literacy/illiteracy statistics tended to give greater prominence to statistics of illiteracy. The reasons for this are explained in Smyth (2006, pp. 2–3).

Commission in 1948 that literacy should be defined in national censuses as “the ability to read and write a simple message in any language” (Smyth, 2006, p. 11).

The recommendation included suggestions of possible methods of measurement:

To determine the number of literates and illiterates, any of the following methods could be used:

1. Ask a question or questions pertinent to the definition above, in a complete census or sample survey of the population.
2. Use a standardized test of literacy in a special survey. The method could be used to verify data obtained by other means or to correct bias in other returns.
3. When none of these is possible, prepare estimates based on (a) special censuses or sample surveys on the extent of school enrollment, (b) regular school statistics in relation to demographic data, or (c) data on educational attainment of the population (UNESCO, 1958, para. 2).

During the period from 1950 to 2000, UNESCO’s international literacy/illiteracy statistics were mostly derived from national censuses (Smyth, 2006).

The classification set out in the 1958 Recommendation emphasized the sex and age breakdown of the literate/illiterate populations:

The population aged 10 years and over should be classified into two groups: literates and illiterates.

Each of these groups should be classified by sex and by age in the following groups: 10–14, 15–19, 20–24, 25–34, 45–54, 55–64, 65 years and over.

Additional classifications should be made, where appropriate, for:

1. Urban and rural populations.
2. Such ethnic groups as are usually distinguished within a State for statistical purposes.
3. Social groups. (UNESCO, 1958, para. 3–5)

Under Statistics on the Educational Attainment of the Population, the recommendation presented a single definition: “The following definition should be used for statistical purposes: The *educational attainment* of a person is the highest grade or level of education completed by the person in the educational system of his/her own or some other state” (UNESCO, 1958, para 6).

The term *grade* was defined elsewhere in the recommendation under Statistics of Educational Institutions, as I soon show. The term *level* was not defined as such

in the recommendation, but four different levels of education were described under the recommendation's classification of Statistics of Educational Institutions.

As in the case of literacy/illiteracy, the recommendation suggested possible methods of measurement:

To measure the educational attainment of the population, the following methods could be used:

- (a) Ask a question or questions pertinent to the definition given above, at a complete census or sample survey of the population.
- (b) Where this is impossible, prepare estimates based on: (i) data from previous censuses or surveys; (ii) records over a number of years of school enrolment, of examinations, of school-leaving certificates, and of degrees or diplomas granted. (UNESCO, 1958, para. 7)

During the period from 1950 to 2000, UNESCO's international statistics on the educational attainment of the population were mostly derived from national censuses.

Under Statistics of Educational Institutions, the 1958 Recommendation's definitions were in most cases revisions of those put forward by the 1951 Expert Committee:

The following definitions should be used for statistical purposes:

- (a) A *pupil (student)* is a person enrolled in a school for systematic instruction at any level of education. (i) A *full-time pupil (student)* is one who is enrolled for full-time education for a substantial period of time. (ii) A *part-time pupil (student)* is one who is not a full-time pupil (student).
- (b) A *teacher* is a person directly engaged in instructing a group of pupils (students). Heads of educational institutions, supervisory and other personnel should be counted as teachers only when they have regular teaching functions. (i) A *full-time teacher* is a person engaged in teaching for a number of hours customarily regarded as full-time at the particular level of education in each State. (ii) A *part-time teacher* is one who is not a full-time teacher.
- (c) A *grade* is a stage of instruction usually covered in the course of a school year.
- (d) A *class* is a group of pupils (students) who are usually instructed together during a school term by a teacher or by several teachers.
- (e) A *school (educational institution)* is a group of pupils (students) of one or more grades organized to receive instruction of a given type and level under one teacher, or under more than one teacher and with an immediate head. (i) A *public school* is a school operated by a public authority (national, federal State or provincial, or local) whatever the origin of its financial resources.



- (ii) A *private school* is a school not operated by a public authority, whether or not it receives financial support from such authorities. Private schools may be defined as *aided* or *non-aided*, respectively, according as they derive or do not derive financial support from public authorities.
- (f) The *compulsory school age* population is the total population between the age limits of compulsory full-time education. (UNESCO, 1958, para. 11)

International tables of the percentage of the “compulsory school-age population” enrolled in school evidently needed to include information on the relevant age range for each country if the percentages for the different countries were to be meaningfully compared. The recommendation’s definitions of *public* and *private* and *aided* and *nonaided* schools drew a clearer distinction than had been drawn by the 1951 Expert Committee between responsibility for the operation of the school and the sources of the school’s finance.

The classification put forward for Statistics of Educational Institutions differed significantly from the earlier proposals of the 1951 Expert Committee. The latter had distinguished between four categories of education: education classified by level, teacher education, special education, and supplementary education (including adult education), with the first category basically referring to regular schools and university/higher education institutions, which were then classified by level of education. The 1958 Recommendation distinguished simply between education that is usually classified by level and education which is not usually classified by level, with the former broken down into three levels and a level “preceding the first level”:

Education should be classified as far as possible by level as follows:

- (a) Education *preceding the first level*, which provides education for children who are not old enough to enter school at the first level (e.g., at nursery school, kindergarten, infant school).
- (b) Education at the *first level*, of which the main function is to provide basic instruction in the tools of learning (e.g., at elementary school, primary school).
- (c) Education at the *second level*, based upon at least four years previous instruction at the first level, and providing general or specialized instruction, or both (e.g., at middle school, secondary school, high school, vocational school, teacher training school at this level).
- (d) Education at the *third level*, which requires, as a minimum condition of admission, the successful completion of education at the second level, or evidence of an equivalent level of knowledge (e.g., at university, teachers’ college, higher professional school).

Education which is not usually classified by level should be placed under one of the following headings:

- (a) *Special education*, covering all general or vocational education given to children who are physically handicapped, mentally handicapped, socially maladjusted or are in other special categories.
- (b) *Other education*. (UNESCO, 1958, para. 12–13)

This classification resolved the problem of the numbering of the levels that had arisen with the 1951 Expert Committee's classification. However, it did not specify the unit of classification. For the 1951 Expert Committee, the unit of classification was the school or educational institution. For the 1958 Recommendation there was no classification unit as such; there simply were different levels of education. In practice, after the recommendation was adopted, most countries probably took the school or educational institution as the unit of classification and then placed their schools and educational institutions as appropriate into one or another of the recommendation's levels of education. There would have been a difficulty in respect of schools or institutions that provided education at more than one level, but this difficulty would have arisen in only a minority of cases. The difficulty was later resolved by ISCED, which put forward the "educational programme" as the unit of classification.

As in the case of the 1951 Expert Committee's classification, though with a revised numbering, the 1958 Recommendation benchmarked the levels on elementary/primary education, which became the first level under the revised numbering. Thus, although avoiding a formal definition of the term *level*, the intergovernmental committee that drew up the recommendation, like the 1951 Expert Committee previously, basically conceived the "levels" as steps on a ladder, in effect successive stages in a process of instruction running from infant or nursery school ("education preceding the first level") up to university. This conception of levels of education has essentially been retained in UNESCO's educational statistics down to the present day. The "levels" terminology, as noted earlier, was artificial in the literal sense of having been "contrived," but it may be more readily understood as metaphorical, the metaphor being that of a ladder of indeterminate length with a certain number of steps.

For both the second and third levels, the recommendation added a classification by type of education:

Where possible, education at the second level should be sub-divided by type as follows:

- (a) *General education*, which does not aim at preparing pupils directly for a given trade or occupation. Where appropriate, general education should be further subdivided as follows: (i) *lower stage*, in which general instruction is given, with orientation of pupils according to interests and aptitudes (e.g., at junior middle school, junior secondary school, junior high school); education at this

stage may lead to various types of instruction at a higher stage; (ii) *higher stage*, in which some differentiation is provided in the types of instruction according to the interests and aptitudes of the pupils (e.g., at senior middle school, senior secondary school, senior high school).

- (b) *Vocational education*, which aims at preparing the pupils directly for a trade or occupation other than teaching. Where appropriate, vocational education should be further subdivided as follows: (i) education which is mainly *practical*; (ii) education which is mainly *technical and scientific*.
- (c) *Teacher training*, which aims at preparing pupils directly for teaching.

Education at the third level should, as far as possible, be classified by type as follows:

- (a) Education at universities and equivalent institutions leading to an academic degree;
- (b) Teacher education at non-university institutions;
- (c) Other education at non-university institutions. (UNESCO, 1958, para. 14–15)

As in the case of the levels themselves, the application of the 1958 Recommendation's classification by type of education was problematic for schools and educational institutions that provided for more than one type of education, especially if the school or educational institution was taken as the unit of classification. Thus, up until the late 1970s, when ISCED overcame this problem by introducing the educational programme as the unit of classification, it is unclear what practices were actually followed by countries when reporting their second- and third-level educational statistics to UNESCO, although it is likely that they reported according to the type of education that mainly characterized the school or educational institution in question. For example, countries with general secondary/high schools that provided various forms of "vocational" education besides "general" education, as in the United States, would typically have reported the student enrollments in these schools as enrolments in general education.

The recommendation's provisions under the fourth and last heading, Statistics of Educational Finance, basically concerned the classification of financial receipts and expenditures of educational institutions. The provisions called for states to present tabulations of financial receipts and expenditures "corresponding as near as possible" to the classification set out earlier for educational institutions, with receipts broken down into public and private sources and expenditures broken down by recurring and capital expenditure (UNESCO, 1958, para. 19).

## THE NEED FOR ISCED

In setting out for the first time a broadly acceptable conceptual framework for the compilation of international educational statistics, the 1958 Recommendation



removed much of the uncertainty that hitherto had surrounded UNESCO's role in compiling such statistics. The provisions of the recommendation were quickly absorbed into the organization's analytical work as well as its data collection operations. On the analytical side, in the World Survey of Education, for example, international trends in school enrollments would henceforth be monitored by reference to "level of education":

For the first time since the adoption by the General Conference of UNESCO, in 1958, of the Recommendation concerning the International Standardization of Educational Statistics, an attempt will be made to present a world summary of school enrolment by continents and regions and by level of education. For better comparability, this summary omits all figures relating to education preceding the first level (pre-primary education), as well as other types of education not classifiable by level (notably figures relating to special education, and of various types of adult education). Nevertheless, the presentation of enrolment data according to the three principal levels does involve certain arbitrary choices regarding the classification of different types of schools. For example, higher primary, intermediate or middle schools have generally been included under the second level of education; vocational and teacher training schools are for the most part included under the second level, except for those technical schools and teacher training colleges requiring, as a condition of admission, the completion of education at the second level; under the third level are included all universities and other institutions of higher education, as well as technical, teacher training and other types of schools above the level of secondary education. (UNESCO, 1961b, p. 26)

The difficulty of classifying enrollment by level of education evidently was due to the continuing use of the school or educational institution as the unit of classification.

On the operational side, the recommendation enabled UNESCO to institute a system of standardized national reporting utilizing annual statistical questionnaires. To assist countries in preparing their responses to the questionnaires, a *Manual of Educational Statistics* was published in 1961 aiming "to explain [the recommendation's] suggestions concerning definitions, classifications and tabulations of educational statistics" (UNESCO, 1961a, p. 8). With standardized national reporting established on an annual basis, the *Statistical Yearbook* (started in 1963) became UNESCO's main vehicle for the publication of international educational statistics.<sup>15</sup>

The organization's attention was soon drawn toward questions concerning the uses of educational statistics. Users' needs in the late 1950s and early 1960s were

---

<sup>15</sup>The *Statistical Yearbook* replaced an earlier series, *Basic Facts and Figures* (1952–1962), which presented selected educational statistics based largely on data provided by national publications (see UNESCO, 1963–2000).

evolving rapidly as education increasingly came to be recognized by national policymakers as a critical factor in economic and social development. Regional indicative plans for the development of education in Asia and Africa were adopted by Conferences of States convened by UNESCO in Karachi (1960) and by UNESCO jointly with the United Nations Economic Commission for Africa in Addis Ababa (1961), with the aim broadly of achieving universal primary education in these regions by 1980. In the industrial countries, the “Sputnik shock” (1957) brought the question of a possible shortage of scientific and technological manpower to the forefront of national security concerns. In the growing number of developing countries that had recently achieved (or were about to achieve) independence, scarcities of “high-level manpower” in particular had come to be seen as a major constraint on development. There was a surge of academic interest in, and research into the role of education in economic and social development among social scientists (economists, sociologists and anthropologists).<sup>16</sup>

The role of statistics in educational planning and policymaking had largely remained in the background during the long process that led up to the adoption of the 1958 Recommendation, but it came to the fore in the early 1960s as a result of the intense interest, in both industrial and developing countries, in the relationship between education and the economy’s and society’s needs for highly qualified manpower. As a leading observer noted at the time,

What is new about educational planning in our own day is the degree to which more and more countries are subordinating the expansion of the educational system to the prospective demand of government and industry for highly qualified manpower, a prospective demand which is forecast with ever more sophisticated techniques. (Blaug, 1966, p. 71)

In 1963, barely 5 years after the adoption of the 1958 Recommendation, the head of UNESCO’s Statistics Division had come around to the view that “there is a need for work on an educational classification system capable of cross-classification with occupational and industrial classifications, one outcome of which would be the provision of a scheme helpful to manpower and educational planners” (Holmes & Robinson, 1963 p. 32). This view contained the germ of what was eventually to become ISCED. There already existed an International Standard Classification of Occupations (ISCO) drawn up by the ILO (1958) and an International Standard Industrial Classification of all Economic Activities drawn up by the United Nations (1958). At that time, in countries that were engaged in manpower planning, the economy’s “manpower requirements” were typically derived from projections

---

<sup>16</sup>For a comprehensive annotated bibliography of the post-Second World War international literature on education and economic and social development, including the literature relating to “manpower planning and forecasting,” see Blaug (1966).

of future economic output and employment by industry sector broken down by occupational category. What the head of UNESCO's Statistics Division particularly had in mind was the need for a classification of education that could be applied to occupations, thus enabling projected manpower requirements (i.e., employment by occupational category) to be translated into the numbers of graduates that the educational system would need to produce in different fields of study. For that purpose the educational classification set out in the 1958 Recommendation was clearly inadequate, as it recognized only three broad levels of education and did not identify fields of study other than a partition at the second level between general and vocational.

In effect, educational statistics were being called upon to play a new role that had not been foreseen when the 1958 Recommendation was drawn up:

It is needless to elaborate on the change in the role of educational statistics. With the new emphasis on educational planning and the use of technical skills, the link between industry's manpower needs and vocational training, and the growing recognition of the crucial importance of an adequate and well balanced educational system for the economic development of a country, a completely new approach to educational statistics has developed. Planners and economists have discovered serious gaps in the existing programmes of data collection in the field of education and science. As one of our field experts has commented: 'The educational system can from a certain point of view be considered as an enterprise, supposed to produce the skilled manpower for the country concerned. This enterprise absorbs enormous, and successively increasing financial resources, but, contrary to other enterprises of a purely economic character, it is one which has practically no meaningful and detailed statistics to measure its productivity'. It should be remembered that this is an evaluation which is not limited to underdeveloped areas but holds true for most of the developed countries. (Kappel, 1966, pp. 662-663)<sup>17</sup>

UNESCO's Statistical Division therefore embarked on the design of a new educational classification:

In cooperation with the International Labour Organization, an international system of educational classification is under study and development and is being designed so that it can be cross-classified with the international system of classification of occupations. It is expected that this tool will prove very valuable for purposes of educational planning, especially in planning educational output in manpower and occupational terms. (Kappel, 1965, p. 663)

---

<sup>17</sup>Kappel was at that time the director of UNESCO's Office of Statistics, which the Division of Statistics had then become. His paper is reproduced in the ISI Compendium cited in footnote 2.



## THE PURPOSES, SCOPE, AND CONTENTS OF ISCED

A decade was to elapse before the new educational classification was finalized. This was only slightly longer than the period required for drawing up the 1958 Recommendation, in both cases the slowness being due to the need for extensive international consultations at each stage of the drafting process.<sup>18</sup> In the case of ISCED, following a preliminary discussion by a joint UNESCO–ILO working group in May 1966,<sup>19</sup> the successive stages of the drafting process are marked by a series of expert group meetings and international consultations held every two years from 1968 up until 1974 when a consensus was reached on a draft that could be submitted to the International Conference on Education (ICE) for examination at its 1975 session in Geneva. Following its approval by the ICE, ISCED was then incorporated into a Revised Recommendation concerning the International Standardization of Educational Statistics, which UNESCO's General Conference adopted in December 1978.

From the beginning there was general agreement that the purposes of the new classification were to facilitate international compilation and comparison of educational statistics as such and their use in conjunction with manpower and other economic statistics. Although the first of these purposes had also been the aim of the 1958 Recommendation, the second was new and was largely to determine the way in which ISCED would be developed. In the manpower field the closest available classification was the ISCO, which had been conceived very broadly as a comprehensive classification of the world of work essentially defined as the universe of occupations. In a similar fashion it was decided that ISCED would be a comprehensive classification of education defined as the universe of “organized and sustained communication designed to bring about learning”:

For the purposes of ISCED, then, *education* is taken to comprise *organized and sustained communication designed to bring about learning*. *Communication* requires a relationship between two or more persons involving the transfer of information. *Organized* is intended to mean planned in a pattern or sequence with established aims or curricula. It involves an educational agency which organizes the learning situation and/or teachers who are employed (including unpaid volunteers) to consciously organize the communication. *Sustained* is intended to mean that the learning experience has the elements of duration and continuity. *Learning* is taken as any change

---

<sup>18</sup> An historical account of the process of drawing up ISCED is provided in UNESCO (1992b).

<sup>19</sup> The first draft of the proposed new educational classification was presented to the working group by a consultant, Mr. N. L. McKellar, of the Dominion Bureau of Statistics, Canada, who was also closely associated with an ongoing revision of the ISCO then being carried out by ILO. Copies of McKellar's first draft and a revised draft sent to a number of countries for comment in September 1966 have not survived—at least, a search for them in UNESCO's archives in 1992 was unsuccessful (UNESCO, 1992b, pp. 3–4).

in behaviour, information, knowledge, understanding, attitudes, skills or capabilities which can be retained and cannot be ascribed to physical growth or to the development of inherited behaviour patterns. Included in this scope, therefore, are activities that in some countries and in some languages may not usually be described as “education”, but rather as “training” or as “cultural development”. Excluded, however, are types of communication that are not designed to bring about learning, or that are not planned in a pattern or sequence with established aims. Thus, all education involves learning, but some forms of learning are not regarded as education. Leisure-time activities such as recreation, sports, and tourism which are not designed to bring about learning and which do not involve an educational agency are excluded. “Self-directed learning”, “family and socially-directed learning” and “random learning” are excluded because they involve no organized agency or teacher (in the above sense), as are isolated events involving no sustained educational activity, such as one or two public lectures, conferences or meetings; entertainment; information, advertising and selling programmes; other social and corporate activities, such as meetings of clubs or associations or work camps. (UNESCO, 1976, pp. 2–3)

The “universe of education” thus defined was then partitioned into four major categories of education:

Within the framework of ISCED, the universe of education will include several categories which also need to be defined. Two major categories are as follows:

- *Regular school and university education*: This is used here to describe the system that provides a ‘ladder’ by which children and young people may progress from primary schools through universities (although many drop out on the way). It is designed and intended for children and young people, generally beginning at age five to seven up to the early twenties (although in some circumstances other students are accommodated along with their younger colleagues).
- *Adult education*: This is used here to describe out-of-school education, which provides education for people who are not in the regular school and university system and who are generally fifteen or older (although in some circumstances, younger students are accommodated along with their older colleagues).

Two other major categories that should be distinguished for statistical purposes are:

- *Formal education*: i.e. education in which students are enrolled or registered, regardless of the mode of teaching used; i.e. it includes an educational series transmitted by radio or television if the listeners are registered.
- *Non-formal education*: i.e. education in which students or ‘clients’ are not enrolled or registered.

In this sense, all regular school and university education is essentially *formal* in that students are enrolled. Adult education, however, can be *formal* or *non-formal*, and this distinction is useful statistically in that measurement of participation by students

or clients presents particular problems in the absence of enrolment or registration. (UNESCO, 1976, p. 19)

Thus, ISCED dropped the 1958 Recommendation's partitioning of education into two broad categories, "education that is usually classified by level" and "education that is not usually classified by level." In ISCED, most forms of education, whether regular school and university education or adult education, could be classified by level. ISCED took over the notion of level from the 1958 Recommendation but divided each of the latter's second- and third-"level categories" into a first and a second "stage," thus in effect adding two more "level categories" to those set out in the 1958 Recommendation:<sup>20</sup>

- 0 Education preceding the first level
- 1 Education at the first level
- 2 Education at the second level, first stage
- 3 Education at the second level, second stage
- 5 Education at the third level, first stage, of the type that leads to an award not equivalent to a first university degree
- 6 Education at the third level, first stage, of the type that leads to a first university degree or equivalent
- 7 Education at the third level, second stage, of the type that leads to a postgraduate university degree or equivalent
- 9 Education not definable by level (UNESCO, 1976, p. 5).

The most significant innovation in ISCED was to conceive of "education," whether regular school and university education or adult education, as made up of "units" of education in the form of "courses" and "programs" that could be classified by level category and aggregated within each level category into "program groups" and "fields" of subject-matter content, which in turn could be correlated—more or less depending on the content—with occupations or groups of occupations. In this way ISCED provided for the much sought-after link between education and manpower planning. The course and the program were defined as follows:

A course ... is taken to be a planned series of learning experiences in a particular range of subject matter or skills offered by a sponsoring agency and undertaken by one or more students.

---

<sup>20</sup>"A final position 'X No Education' can be provided as required, e.g. when obtaining statistics of the stock of educated people from an enumeration of the population of an area as in a population census. Such a category is not needed for statistics of current educational operations" (UNESCO, 1976, p. 5). The digits 4 and 8 were not utilized, and no reasons were given in the ISCED documentation, although it has since been thought that ISCED's architects intentionally left gaps so as to allow for the possibility of inserting new level categories in future revisions of ISCED.



A programme . . . is taken to be a selection of one or more courses or a combination of courses usually chosen from a syllabus, a calendar, or a list. Such a programme may consist of one or a few courses in a specific field or, more commonly, of a number of courses most of which will be classified within a specific field but some may be classified in other fields. Each programme has an expressed or implied aim, such as qualification for more advanced study, qualification for an occupation or range of occupations, or solely an increase in knowledge or understanding. (UNESCO, 1976, p. 4).

ISCED then classified programs by subject-matter content (UNESCO, 1976, pp. 4, 8–15, 30), which had not previously been attempted except in so far as the 1958 Recommendation drew a distinction between “general education” and “vocational education” at the second level. Programs that could be considered as related in terms of level and major subject matter content were taken to constitute “program groups,” which were then aggregated into subject-matter “fields.” A five-digit coding system was developed with the first digit taken to represent the level, the second and third digits the field within the level, and the fourth and fifth digits the program group within the field.<sup>21</sup>

The most detailed categories in ISCED are groups of programmes that are related in terms of level and subject-matter content, e.g. programmes in history at a given level (each such programme group being identified by a five-digit code number). Programme groups are further aggregated into fields composed of programme groups related to the same general subject matter within a level category, e.g. humanities programmes at a given level (each field being identified by a three-digit code number). Fields and their constituent programme groups are designated within *level categories which, as their name implies, are categories representing broad steps of educational progress from very elementary to more complicated learning experience* [italics added] (each level category being identified by a one-digit code number). ISCED is, therefore, a three-stage classification system containing groups in a hierarchical arrangement from very broad level categories to broad subject-matter fields to narrower subject-matter programme groups (the programmes constituting programme groups are composed of courses which represent the smallest educational units recognized in the ISCED system of definitions, but courses are not specified separately in the classification system and are not assigned code numbers). (UNESCO, 1976, p. 21)

The determining factor for placing a particular program or group of programs at a given level was taken to be “the minimum prior education required to take advantage of the programme” (UNESCO, 1976, pp. 6–7, 25). For a program at

---

<sup>21</sup>It was recognized in ISCED that “some fields do not exist at every level, e.g. law and jurisprudence programmes are not found at level categories 2 or 3 while literacy programmes occur only at level category 1” (UNESCO, 1976, p. 13).

the second level, first stage, for example, the minimum prior education would normally be the completion of a program at the first level. Likewise, the minimum prior education required to take advantage of a program at the third level, first stage (whether Level Category 5 or Level Category 6), would normally be the completion of a program at the second level, second stage. Thus, like the 1958 Recommendation, ISCED in effect “anchored” or benchmarked its classification of the level categories on the education provided at the first level. However, ISCED’s architects went further than their predecessors by claiming that the cross-national comparability of the classification was ensured by the existence of a “core of education for young people in most countries [that] can be expressed as a sequence of stages, each being encompassed in a number of years of full-time education”<sup>22</sup> (UNESCO, 1976, p. 7).

Although the time usually spent by students in particular stages varies from country to country, the overall sequence is found to be quite uniform and the total time (i.e., full-time equivalent) spent by a typical student from original school entry to university graduation is quite consistent around the world. Thus, if the disparate stages in national systems imposed by the national pattern of educational institutions can be ignored, it is found that an internationally applicable set of ISCED level categories for the universal educational core can be described very briefly as follows:

- 0 *Education preceding the first level*, where it is provided, usually begins at age three, four or five (sometimes earlier) and lasts from one to three years.
- 1 *Education at the first level* usually begins, therefore, at age 5, 6 or 7 and lasts for about five or six years.
- 2 *Education at the second level, first stage*, begins at about age 11 or 12 and lasts for about three years.
- 3 *Education at the second level, second stage*, begins at about age 14 or 15 and lasts for about three years.
- 5 *Education at the third level, first stage, of the type that leads to an award not equivalent to a first university degree*, begins at about age 17 or 18 and lasts for about three years. Thus, at about age 20 or 21, students who have progressed through the regular school system to complete these programmes are usually ready to enter employment.
- 6 *Education at the third level, first stage, of the type that leads to a first university degree or equivalent*, also begins at about age 17 or 18 and lasts for about four years. Thus, students who have progressed through the school system to complete their first degree are usually ready for employment or for postgraduate study at about age 21 or 22.

---

<sup>22</sup>“Despite its known variability, a full-time year at school successfully completed is the most objective unit of education available as an international ‘yardstick’” (UNESCO, 1976, p. 6).

- 7 *Education at the third level, second stage*, of the type that leads to a post-graduate university degree or equivalent, includes all education beyond level 6.

The above is a formalized sketch of the core intended to identify it for the purpose of international definition. The summary merely provides a scale or measuring rod that can be used to identify corresponding stages in any national system. The core, however, does not contain all the educational programmes that can be classified by level. Many programmes of out-of-school and vocational education or training (often lumped under the heading of adult education) deal with subject matter requiring previous formal education on the part of those who undertake them. (UNESCO, 1976, pp. 25–26)

Because the determining factor for placing a particular program or group of programs at a given level was “the minimum prior education required to take advantage of the programme,” the existence of the “core” did not preclude the possibility that in practice many programs at a given level, particularly those of a vocational/occupational orientation, could be terminal and of shorter or longer duration than the number of years normally required in programs leading to admission at the next level.

ISCED’s architects were aware that the full gamut of levels, fields, and program groups would rarely, if ever, be applied in practice. It was assumed that users would apply ISCED as appropriate for the purposes of the particular survey that they planned to undertake:

Surveys of some types . . . will need even more detail than is provided by the ISCED programme groups, e.g. special surveys of higher education collecting information on detailed subject categories. ISCED fields and programme groups can be subdivided and the blank spaces in the three-digit and five-digit code system used. Other surveys or tabulations may require levels of detail falling between the steps in the ISCED hierarchy. A likely example of this kind will be statistical analyses of data derived from sample surveys and requiring cross-classifications of educational factors with personal or non-educational characteristics. The eight ISCED ‘level’ categories are likely to be too broad for meaningful analysis, while the ‘levels’ and ‘fields’ comprising some 100 groups may be too detailed for tables involving cross-classification [e.g. with occupations or groups of occupations]. An intermediate grouping having something less than 20 groups like the following [three-digit groups] could be useful:

1. Level 0 – Education preceding the first level
2. Level 1 – General education at the first level except literacy programmes (101)
3. Level 1 – Other programmes of education at the first level (126, 134, 150, 152, 162, 166, 178, 189)
4. Level 1 – Literacy programmes (108)
5. Level 2 – Programmes of general education (201)



6. Level 2 – Teacher-training programmes (214)
7. Level 2 – Other programmes of education at the second level, first stage (226, 234, 250, 252, 262, 266, 270, 278, 289)
8. Level 3 – Programmes of general education (301)

[... and so on]. (UNESCO, 1976, pp. 31–32)

Such “intermediate” groupings evidently could be useful for the purposes of education and manpower planning at the national level, but at the international level their usefulness would depend on the comparability of the levels and fields across countries. For ISCED’s architects, confident of having devised a suitable framework for assembling internationally comparable data, the challenge in most countries would be to obtain data in the degree of detail needed for the purposes in question:

[ISCED] is designed for assembling data on current educational phenomena such as enrolment, teaching staff and finances as well as for statistics of the “stock” of educated people as obtained, for example, by a census of population. In this sense it is a multi-purpose system within which comparable data can be assembled on various features of educational systems and processes. Of course, it is not feasible to assemble data on all such features to the same degree of detail because of the different units to which the data relate. Enrolment figures for example, which relate to individuals enrolled in particular programmes can usually be reported in more detailed categories than can information on teachers, many of whom are involved in a number of programmes. Some kinds of financial information such as assets, liabilities and fixed capital employed, are usually available only for units like institutions (or groups of institutions under common management, e.g. a local educational authority). “Stock” data as obtained from a population census are usually collected only in terms of the “highest educational level or grade attained” by each individual. (UNESCO, 1976, p. 1)

ISCED was not put forward as a replacement for national classification systems, where such systems were already in place. The expectation was that countries which were already using a comprehensive national classification system would map this system onto ISCED when reporting their statistics internationally:

UNESCO does not expect that those countries now using a comprehensive national classification of education will replace it with ISCED for national compilations. On the contrary, the special requirements of countries for nationally-based classifications are understood and the value of national classifications will be enhanced when, being designed to achieve comparability with ISCED, they can be used to provide internationally comparable data in addition to statistics reflecting particular national patterns of education. Many countries, however, have not yet developed comprehensive national classifications of education, and they may choose to adopt ISCED as it

stands or modified to suit national conditions. Any modifications introduced should be carefully designed to ensure that the resulting data can be rearranged into the ISCED pattern for international reporting. (UNESCO, 1976, p. 1)

## EPILOGUE

The education-manpower planning rationale that had originally led to the formulation of ISCED was already beginning to fall out of favor among educational policymakers and planners when ISCED was finally incorporated in a Revised Recommendation concerning the International Standardization of Educational Statistics (1978). This was largely the result of a broader worldwide shift in approaches to national development, away from state planning and toward the market economy, a movement that was to gather pace in the 1980s. In the human capital view of education and development that eventually prevailed, the link between education and occupation, though still relevant, was less important than the link between education and earnings.<sup>23</sup> Among economists, emphasis came to be placed on the "rate of return" on investment in education, and few if any attempts were made to apply the full panoply of ISCED-level categories, fields, and program groups in education-manpower planning exercises.

The main features of ISCED, notably the level categories and broad groupings of fields, were progressively incorporated into UNESCO's educational statistics. But as had happened after the adoption of the 1958 Recommendation, the priority needs of educational statistics users continued to evolve. By the end of the 1980s, attention was increasingly focused on the internal efficiency of education systems and learning outcomes. In 1989 OECD's Centre for Educational Research and Innovation initiated the Indicators of National Education Systems Project, a long-term project for the development of comparative international indicators of education system inputs, processes, and outcomes (OECD, 1991a, 1991b). In the following year, the World Conference on Education for All (Jomtien, Thailand, 1990) was to place particular stress on "learning achievement."<sup>24</sup>

---

<sup>23</sup>A comprehensive historical review of the worldwide manpower planning movement of the 1950s and 1960s has never been written. A useful critique of education-manpower planning from a "human capital" point of view containing many references is provided by Psacharopoulos and Woodhall (1985).

<sup>24</sup>"Article 4: *Focusing on Learning Achievement*. Whether or not expanded educational opportunities will translate into meaningful development—for an individual or for society—depends ultimately on whether people actually learn as a result of those opportunities, i.e. whether they incorporate useful knowledge, reasoning ability, skills, and values. The focus of basic education must, therefore, be on actual learning acquisition and outcome, rather than exclusively upon enrolment, continued participation in organized programmes and completion of certification requirements. Active and participatory approaches are particularly valuable in assuring learning acquisition and allowing learners to reach their fullest potential. It is, therefore, necessary to define acceptable levels of learning acquisition for

Classification issues were soon of major concern to OECD, and ISCED came under close scrutiny. The majority of the issues concerned the treatment of program duration. Attention focused particularly on the classification of programs of postcompulsory education and higher education, that is, ISCED-level categories 3, 5, 6, and 7. In most OECD countries there existed a variety of educational programs for young persons who had completed their compulsory education, but "compulsory education" was not an ISCED-level category, and its duration varied from country to country, in some cases coinciding with the Level Categories 1 and 2, and in other cases extending into Level Category 3 (Education at the second level, second stage). Many postcompulsory education programs of a vocational/occupation-oriented character did not prepare students for access to education at the third level, first stage (Level Categories 5 and 6), and in some cases had a shorter duration than postcompulsory education programs that did prepare students for entry to Level Categories 5 and 6. It was unclear in terms of the ISCED "core" hierarchy of years of education, and the ISCED assumption of a "minimum prior education required to take advantage" of a Level Category 5 or 6 program, whether all postcompulsory education programs not classifiable in Level Categories 5 or 6 should be classified in the same level category.<sup>25</sup>

Similar problems arose in respect of the classification of programs at the third level, first and second stages (Level Categories 5, 6, and 7), mainly because of national differences in the duration of both nondegree and first-degree programs.<sup>26</sup> For example, a nondegree program and a degree program of the same duration would be classified by ISCED in different level categories (5 and 6, respectively). Moreover, some countries with long-duration first-degree programs classified these programs in Level Category 7 (Education at the third level, second stage, of the type that leads to a postgraduate degree or its equivalent).

In 1992 UNESCO convened a Meeting of Experts on Education Indicators and the ISCED for the purpose of reviewing these and other issues, for example, the treatment of distance education and adult and out-of-school education, which had arisen with the application of ISCED (UNESCO, 1992c). Several countries at UNESCO's General Conference in November to December 1993 called for UNESCO to undertake a revision of ISCED. In response, the Director-General set up a task force, including the participation of representatives from OECDL and ILO and EUROSTAT charged with drawing up a draft revised version for presentation to the ICE at the latter's 1996 session in Geneva. As approved by the ICE, the revised

---

educational programmes and to improve and apply systems of assessing learning achievement" (*World Declaration on Education for All and Framework for Action to Meet Basic Learning Needs*, 1990).

<sup>25</sup>An extensive discussion of the issues relating to the classification of vocational/occupation-oriented programs is provided in UNESCO (1992d).

<sup>26</sup>An extensive discussion of the issues relating to the classification of third-level programs is provided in UNESCO (1992a).



version of ISCED, incorporating extensive guidance on classification criteria for the various level categories, was adopted by UNESCO's General Conference in December 1997.

ISCED 1997 is still basically the "artificial terminology" of three broad "levels" of education as first presented in the recommendation concerning the International Standardization of Educational Statistics (1958). Whether this terminology satisfactorily overcomes the problem of international comparability in educational statistics depends to a large extent on what the "level categories" are taken to mean. If the meaning attributed to them in ISCED is taken ("broad steps of educational progress from elementary to more complicated learning experience"), then there arises the question of whether the "complexity" of the "learning experience" at a given level is comparable across countries. The problem still remains if the level categories are assumed, as in ISCED, to be broadly correlated with years of schooling. Some economists' estimates of the stock of human capital in different countries, which ultimately are based on UNESCO's international educational statistics, depend on the assumption that a year's schooling in one country can be taken as comparable to a year's schooling in any other.<sup>27</sup>

They also depend on the assumption that a year's schooling can be taken as a proxy measure of "human capital" itself, another artificial terminology in the sense originally suggested by Nicholas Hans.

## REFERENCES

- Blaug, M. (1966). *Economics of education: A selected annotated bibliography*. Oxford, UK: Pergamon.
- Bureau International d'Éducation. (1933). *Le Bureau International d'Éducation en 1932–1933, Quatrième réunion du Conseil* (Publications du Bureau International d'Éducation, No. 31, p. 8). Geneva, Switzerland: Author.
- Hans, N. (1933). Comparative statistics. In Lord E. Percy (Ed.), *The Year Book of education* (pp. lxxxviii–xc). London: Evans Brothers.
- Hanushek, E. A., & Wößmann, L. (2007). *The role of education quality for economic growth* (Policy Research Working Paper No. 4122). Washington, DC: World Bank.
- Holmes, B., & Robinsohn, S. B. (1963). *Relevant data in comparative education* (Report on expert meeting at UNESCO Institute for Education, Hamburg, Germany). Hamburg: UNESCO Institute for Education.
- International Institute of Intellectual Cooperation. (1946). *L'Institut International de Coopération Intellectuelle, 1925–1946*. Paris: Author.
- International Labour Office. (1958). *International Standard Classification of Occupations*. Geneva, Switzerland: Author.
- International Statistical Institute. (1995). *Education statistics: An international compendium 1853–1995*. Voorburg, the Netherlands: Author.
- Kandel, I. L. (Ed.). (1925). *Educational Yearbook of the International Institute of Teachers College, Columbia University, 1924*. New York: Macmillan.

<sup>27</sup>As noted, for example, by Hanushek and Wößmann (2007).

- Kappel, J. W. (1966). The programme of UNESCO in education and science statistics. In *Bulletin of the International Statistical Institute, Proceedings of the 35th Session of the International Statistical Congress (Belgrade, 1965)* (Vol. XLI, Book 2, pp. 661–668). Belgrade: ISI
- Levasseur, E. (1893). *La statistique de l'enseignement primaire*. [Statistics of Primary Education]. In *Compte-Rendu de la Quatrième Session de l'Institut International de Statistique tenue à Chicago du 11 au 15 août 1893*. Volume 2. Voorburg, the Netherlands: ISI.
- March, L. (1928). Rapport au nom de la Commission mixte de la Statistique intellectuelle. Rapports et communications présentés à la XVII<sup>ème</sup> Session de l'Institut International de Statistique, [Report of the Joint Commission on Intellectual Statistics. Reports and Communications Presented at the 17th session of ISI, Cairo] Le Caire 1927–28. In *Bulletin de l'Institut International de Statistique* (Book XXIII, 2nd ed., pp. 609–692).
- Organization for Economic Cooperation and Development. (1991a, 16–18 September). *Education at a glance* (Lugano-Cadro, General Assembly of the INES Project). Paris: OECD, CER/INES.
- Organization for Economic Cooperation and Development. (1991b, 16–18 September). *Handbook on international education indicators* (Lugano-Cadro, General Assembly of the INES Project). Paris: OECD, CER/INES.
- Psacharopoulos, G., & Woodhall, M. (1985). *Education for development*. New York/Oxford: Oxford University Press.
- Report of the Joint Commission on Intellectual Statistics. Reports and Communications presented at the 17th Session [The 17th Session of ISI, Cairo].
- Renoliet, J-J. (1999). *l'UNESCO oubliée: La Société des Nations et la coopération intellectuelle, 1919–1946*. Paris: Publications de la Sorbonne.
- Rossello, P. (1943). *Les précurseurs du Bureau international d'éducation*. [The forerunners of the International Bureau of Education] Geneva, Switzerland: IBE.
- Rossello, P. (1970). Historical note. In *IBE, International Conferences on Public Education: Recommendations 1934–1968* (pp. xi–xxi). Paris/Geneva: UNESCO/IBE.
- Smyth, J. A. (2006). *UNESCO's International Literacy Statistics 1950–2000*. Paper commissioned for the EFA Global Monitoring Report 2006, Literacy for Life (Document No. 2006/ED/EFA/MRT/PI/90). Paris: UNESCO Publishing. Available from UNESCO Web site: <http://unesdoc.unesco.org/images/0014/001461/146185e.pdf>
- UNESCO. (1952). *World survey of education: Handbook of educational organization and statistics*. Paris: Author.
- UNESCO. (1953). *Progress of Literacy in various countries*. Paris: Author.
- UNESCO. (1954). Problems on the standardization of certain aspects of educational statistics. *Bulletin de l'Institut International de Statistique* (Book XXXIV, 3rd ed., pp. 513–517). Rome: ISI.
- UNESCO. (1955). *World survey of education: Handbook of educational organization and statistics* (pp. 47–56). Paris: Author.
- UNESCO. (1958). *Recommendation concerning the international standardization of educational statistics adopted by the General Conference of UNESCO at its tenth session, Paris, 3 December 1958*. Paris: Author.
- UNESCO. (1961a). *Manual of educational statistics* (pp. 26–35). Paris: Author.
- UNESCO. (1961b). *World survey of education III* (p. 26). Paris: Author.
- UNESCO. (1963–2000). *Statistical yearbook*. Paris: Author.
- UNESCO. (1976, March). *International Standard Classification of Education* (Document No. COM.75/WS/27). Paris: Author, Division of Statistics on Education, Office of Statistics.
- UNESCO. (1978). *Revised recommendation concerning the International Standardization of Educational Statistics*. Paris: Author.
- UNESCO. (1992a, June 11–17). *Current issues in the International Comparability of Statistics of Education* (Meeting Document No. ISCED/3, Conference Document No. ST-92/CONF.601/LD.2). Meeting of experts on education indicators and the International Standard Classification of

- Education, Paris. Paris: Author, Education Sector and Division of Statistics. [Document prepared by the Secretariat]
- UNESCO. (1992b, June 11–17). *The early history of the development of the International Standard Classification of Education* (Meeting Document No. ISCED/2, Conference Document No. ST-92/CONF.601/LD.1). Meeting of experts on education indicators and the International Standard Classification of Education, Paris. Paris: Author, Education Sector and Division of Statistics. [Document prepared by Mr W. Hyland, consultant, in collaboration with the Secretariat]
- UNESCO. (1992c, June 11–17). *Final report* (Conference Document No. ST-92/CONF.601/LD.8). Meeting of experts on education indicators and the International Standard Classification of Education, Paris. Paris: Author, Education Sector and Division of Statistics. [Document prepared by the Secretariat]
- UNESCO. (1992d, June 11–17). *Methodological issues and problems concerning the place and classification of technical and vocational education in the International Standard Classification of Education* (Meeting Document No. ISCED/4, Conference Document No. ST-92/CONF.601/LD.3). Meeting of experts on education indicators and the International Standard Classification of Education, Paris. Paris: Author, Education Sector and Division of Statistics. [Document prepared by Ms Christine Mainguet, consultant, in collaboration with the Secretariat]
- UNESCO. (1997). *ISCED 1997* (Document No. BPE-98/WS/1). Paris: Author.
- United Nations. (1958). *International Standard Industrial Classification of all economic activities, revised*. New York: Author.
- Valderama, F. (1995). *A history of UNESCO*. Paris: UNESCO Publishing.
- World Declaration on Education for All and Framework for Action to Meet Basic Learning Needs*. (1990). New York: Inter-Agency Commission (UNDP, UNESCO, UNICEF, World Bank).

## APPENDIX

### *The Preamble of the UNESCO General Conference resolution that adopted the 1958 Recommendation (italics added)*

The General Conference of the United Nations Educational, Scientific and Cultural Organization, meeting in Paris from 4 November to 5 December 1958 at its tenth session,

Considering that Article VIII of the Constitution of the Organization specifies that 'each Member State shall report periodically to the Organization, in a manner to be determined by the General Conference, on its laws, regulations and statistics relating to educational, scientific and cultural life and institutions,'

*Convinced that it is highly desirable that the national authorities responsible for the compilation and reporting of statistics relating to education should be guided by certain standard definitions, classifications and tabulations, in order to improve the international comparability of their data,* Having before it proposals concerning the international standardization of educational statistics which constitute item 15.3.1 of the agenda of the session,

Having decided, at its ninth session, that these proposals should be regulated at the international level by way of a recommendation to Member States, Adopts this third day of December 1958, the present Recommendation:

*The General Conference recommends that Member States should, for the purposes of international reporting, apply the following provisions regarding definitions, classifications and tabulations of statistics relating to education . . .*



# Implementing Free Primary Education Policy in Malawi and Ghana: Equity and Efficiency Analysis

Kazuma Inoue  
*JICA, Niger Ministry of Education*

Moses Oketch  
*Institute of Education, University of London*

Malawi and Ghana are among the numerous Sub-Saharan Africa countries that have in recent years introduced Free Primary Education (FPE) policy as a means to realizing the 2015 Education for All and Millennium Development Goals international targets. The introduction of FPE policy is, however, a huge challenge for any national government that has experienced declining or slow economic growth and heavily relied on charging fees to parents and other sources to finance the education system. It follows, therefore, that the approach taken in implementing the FPE policy has implications for equity and efficiency in the education sector. Malawi and Ghana have differently implemented FPE policy. In this article we assess the impact of the implementation approach taken by each of the two countries on equity and efficiency in their education systems.

The introduction of Free Primary Education (FPE) policy is a big challenge for national governments and international donors in terms of financing. Because the FPE program requires a high budget, national governments and international donors are required to prepare more budgets. In addition, they need to reallocate funds equitably and efficiently by striking a balance between different costs and needs in the education sector.

The discussion in this article is restricted to two particular countries—Malawi and Ghana—that introduced FPE policies in 1994 and 1996, respectively, at an early stage after the establishment of Education for All (EFA). Malawi and Ghana

have some similarities in the context of education and development. Both are former British colonies in Africa with socioeconomic similarities, and consequently they have a common historical background of educational development. They also have similar present educational profiles. They have both been aiming toward similar goals in the context of EFA. In terms of financing, they both increased expenditure on education and are in the process of expanding their educational systems, harmonized with international agencies.

However, the two countries took different approaches in FPE policy implementation. Malawi adopted a policy focusing largely on quantitative expansion with FPE programs and in 1994 eliminated all fees such as tuition fees, uniform, and textbooks (Rose, 2002). On the other hand, Ghana prioritized qualitative improvement and made a FPE policy in 1996, which aimed to reduce a part of the fees/costs for schooling first in 1996 and then to abolish all others by 2005. In fact, Ghana abolished tuition fees in 1996 and other user fees/costs such as uniforms and textbooks in 2005. Ghana first considered qualitative problems, not spending heavily on quantitative expansion. But after a period of time Ghana eliminated all direct costs to seek quantitative expansion (World Bank, 2004a). How did these two different FPE implementation approaches impact on equity and efficiency in the two countries education systems?

The rest of the article is organized as follows. First, we provide an overview of the need for public subsidies of education under various subthemes. Second, we introduce and discuss the concepts of equity and efficiency in education. Third, we provide methodology and situation analysis of FPE policy in Malawi and Ghana. Fourth, we discuss and provide analysis of the impact of FPE implementation approach on equity. Fifth, we discuss and provide analysis of the impact of FPE implementation approach on efficiency. Sixth, we provide the conclusion.

## FINANCING FREE PRIMARY EDUCATION

### *Public Subsidies of Education*

There are three economic arguments that provide the justification for public subsidies to education. The focus of the first argument lies in external benefits of education: Investment in education is important for society, and governments need to avoid underinvestment and to subsidize education from public finance. The second point is the concern for equity and equality of opportunity. If education were managed under the market principle, only those who could afford to pay tuition fees would participate in education. Education itself is a determinant of lifetime income, and inequality of opportunity for education preserves income inequalities from one generation to the next. The third point is that education is believed to be subject to economies of scale. This means that an increased level

of production generates a proportionate saving in costs. Based on this idea, it is more efficient to finance and provide education publicly. These three arguments can be rationales for government subsidies on grounds of both efficiency and equity. These arguments, nevertheless, do not suggest that governments should subsidise all or most of the costs of education. The issues raised by them have more to do with what extent governments should subsidize education—in other words, the optimal balance between public and private financing (Psacharopoulos & Woodhall, 1985).

More specifically, there have been arguments on financing primary education, especially in poor economically underdeveloped nations where governments face competing needs, all of which require urgent attention and prioritization. With regard to the arguments on financing primary education, the focus is especially on whether governments should subsidize all school fees. In the 1960s and the 1970s, charging tuition fees was thought to be inefficient because it discouraged the participation of people for “human capital” and inequitable because it would limit access to the rich. Accordingly, some developing countries raised an ideological commitment for free primary education and actually began to introduce it. However, in the 1980s, because of public financial constraints, some countries began to consider reintroducing school tuition fees. Several analysts (Birdsall, 1982; Mingat & Tan, 1985; Thobani, 1983) maintained that the increase in tuition fees may contribute to both equity and efficiency. The World Bank paper (Psacharopoulos, Tan, & Jimenez, 1986) also proposed that

In general, increased private financing at the primary level is not recommended since it might interfere with universal coverage—a socially desirable goal. But when resource transfers between levels of education and from other sectors are impossible for administrative or political reasons, increased user charges for primary education could increase efficiency within schools, especially if that revenue stays with the school where it was raised. (p. 23)

As a result, the international attention on FPE was weakened. In recent years, however, abolishing school fees has drawn attention again. The trigger is the change of the development paradigm after the 1990s. The World Bank paper (Bentaouet-Kattan & Burnett, 2004) suggests,

Universal primary completion is a top World Bank priority, expressed in the Bank’s commitment to the Millennium Development Goals. The Bank has made abundantly clear in its policy statements that it does not support user fees for tuition in primary education and has in recent years actively supported fee abolition in countries, mainly in Africa, in which fees appear to represent an obstacle to enrolment. (p. 4)



As seen here, the recent paper acknowledges the abolishment of school fees, placing an emphasis on increasing enrollment. Although the justification for abolishing school fees was mainly based on efficiency and equity before the 1980s, it has been placed more on access in recent years. The first reason for this shift is simply because Universal Primary Education (UPE), which aims to increase access, is a top priority among development goals, supported by the concepts of human development, EFA, and Millennium Development Goals (MDGs). Thus, FPE began to draw the attention as a drastic measure for achieving UPE, which has not been achieved since the 1960s, but the impact of FPE on equity and efficiency is still being explored (Bentaouet-Kattan & Burnett, 2004; UNESCO, 2002).

### *Fees for Schooling*

There are various types of fees that private households have to bear even for public schools. The fees or costs for them can be generally divided into two types: direct fees/costs and indirect fees/costs. The direct fees/costs are those spent directly on education—for example, tuition fees, textbook fees/costs, rental payments, compulsory uniforms, PTA dues, and various special fees such as exam fees and contributions to district education boards. The indirect fees/costs are not spent directly on education itself but are unexpectedly required, such as travel fees and loss of work on their firms or their households (UNESCO, 2003). The definition of “Free” Primary Education (in other words, what fees are free) does not have a formal consensus and the interpretation is less straightforward than it might at first seem. FPE, however, can be seen as free for tuition fees, uniform fees, textbooks, and so on in recent contexts (UNESCO, 2002). Hence, even though FPE is introduced, not all fees/costs that pupils and households bear are removed. Bentaouet-Kattan and Burnett (2004) claimed that indirect costs can be an even greater obstacle to schooling than tuition fees. Nevertheless, the alleviation of school fees by FPE removes some of the financial burden of households and encourages increase of access. Also, because some of the financial burden is removed equally for all, although some financial obstacles remain, the degree of opportunity becomes more similar for all financially. Therefore, FPE may be expected to reduce inequality of opportunities by income difference, gender, or geographical region.

### *Provisions for Financing Free Primary Education*

Although FPE subsidizes only parts of fees/costs, the financial burden of governments that finance FPE is huge. In addition, considering that fees make a contribution to overall budgets and to investment for quality of education, simple abolition of fees may have undesired consequences. Bentaouet-Kattan and

Burnett (2004) argued that “adequate measures must therefore be in place to provide equivalent revenues to finance the expenditure previously covered through fee revenue” (p. 5). Moreover, to replace the sources of such revenues, they suggested that national governments should increase expenditure on education by switching spending from other sectors or by increasing revenues. They also claimed to improve the efficiency of education spending, particularly the balance between different education levels and the balance between salaries and other expenditures. Colclough and Al-Samarrai (1998) raised the implications on financing primary education for the achievement of EFA, although not specifically only to financing FPE. They suggested that key factors for sound financing system for EFA are to improve the efficiency of schooling, to reform cost structures so as to reduce unit costs of provision, and to give proper priority to public spending on primary education.

### *Equity and Efficiency*

*The concepts of equity.* Two types of equity are mainly distinguished in the economics literature. These are called “distributional” and “procedural equity” (Musgrave, 1959). Distributional equity refers to the distribution of resources and outcomes such as subsidies, income, benefits received, and educational attainment (Levacic, 2005). However, because people vary by their sex, age, religious belief, interests, needs, culture, and so on, should the distribution also vary to promote equity? With respect to the question, there are two distinctive concepts—horizontal and vertical equity. In the concept of horizontal equity, according to Monk (1990), equity is defined as identical treatment within groups and requires “equal treatment of equals.” Advocates of vertical equity, on the other hand, focus on the differing needs of students and claim that “unequal treatment of unequals” is required to achieve equity. Both of these concepts are thought of as equitable, but attention should be called to groups of people and whether they are equals or unequals, and how they are considered in both equity concepts.

“Procedural equity” focuses on the rules or processes of resource allocation (Levacic, 2005). Because of some conceptual difficulties raised by the distributional equity criteria countered earlier, there has been no consensus on what an equitable distribution of education resources involves (Monk, 1990). However, Wise (1968) emphasized the resource allocation aspects of equality. He defined equality of educational opportunity as existing “when a child’s educational opportunity does not depend upon either his parents’ economic circumstances or his location within the state” (p. 146). This is a negative definition of equality of educational opportunity. Le Grand (1991) stated that “a distribution is equitable if it is the outcome of informed individuals choosing over equal choice sets (p. 87). This gives more emphasis to the relationship with the existence of choice.

Compared to the definition of Wise, however, Le Grand's statement implies that even if uniform subsidies are provided for everyone, that does not necessarily provide equal choice sets. For instance, even if school fees are waived to improve opportunities for all children, the children from poor households tend to face more barriers than others because they may have to earn money to support their family and so still do not have equal choices with others. Thus Le Grand's concept may require positive discrimination with poor students receiving larger grants than students from well-off backgrounds.

*The concepts of efficiency.* There are two different aspects of efficiency in economics. They are called exchange efficiency and production efficiency (McMeekin, 1975).

Exchange efficiency is the efficiency in the exchange or delivery of a given stock of goods and services. This definition seeks the best fit between distribution and needs for the best utility. Monk (1990) discussed exchange efficiency by distinguishing between a given stock of goods and services and a variable level of satisfaction or utility experienced by individuals from those goods and services. He claimed that goods and services should be desirable and contributory to the well-being of people. Moreover, he argued that the distribution of the goods and services among people should contribute to satisfaction or utility. The concept of exchange efficiency is rooted in the utilitarian notion that "the general good is served by maximizing the average level of utility in the society where the average utility is defined as the total utility divided by the number of individuals" (Monk, 1990, p. 4). This notion of utilitarianism provides a base of the idea that when the number of individuals and resources is limited, one of the ways to increase the total level of utility is a suitable combination in the process of exchange.

From the standpoint of policymaking, exchange efficiency is enhanced when the combination of distribution decided by policymakers and individuals' needs is achieved and utilized well (Johnes, 1993). In the education sector, therefore, exchange efficiency encompasses changes in the structure of the educational system and in the number of students in each level until it fits the needs of students and society. This is because human capital skills and knowledge are embedded in individuals and untransformable between students. Hence, the allocation of resources is necessary among educational levels and institutions, adequately matching needs and abilities for good utilization. An example of this is the development of community schools, which may improve the exchange efficiency by facilitating more choices to meet needs (McMahon, 1982).

Production efficiency, in contrast to exchange efficiency, refers to the efficiency in producing goods and services rather than exchanging them. Exchange efficiency assumes that the amount of goods and services is fixed as a supply, so that a particular efficient distribution is required for utility. In contrast, production



efficiency assumes that the goods and services as a output should be produced and if more or better are gained at the same amount of inputs or costs, efficiency would be improved (Monk, 1990).

Production efficiency is distinguished in two dimensions. The first dimension is called “technical efficiency” and focuses on technical aspects in the production process through which inputs are transformed into outputs. This efficiency assesses what particular mix of given inputs the producer should use to maximize output. It is concerned with the quantitative relationship between output and inputs (Mace, 2000).

The second dimension in the production context is known as “allocative efficiency” or “price efficiency.” Price efficiency focuses, on the other hand, on the efficiency as a maximization of the least cost. It analyses the different mixes of inputs that can produce the same outcome (Johnes, 1993). This efficiency is concerned with the relationship between output and the cost of inputs (Mace, 1996). Price efficiency is improved when the same output is produced at a lower cost, or when greater output is produced at the same cost. To give an example of the evaluation of this efficiency, let us suppose that the learning environment in a classroom contributes to a better score in an exam. What method in the learning environment is the least costly? The efficiency is compared with the rates of output and costs of some methods, (e.g., one-to-one interaction, text-based teaching or more sufficient materials with fewer teachers).

## METHODOLOGY

This article aims to achieve the objectives previously stated by reviewing secondary sources such as reports, government documents, academic books, and journals on the state of education in Malawi and Ghana. This study is document-based research. Data on enrollment trends, expenditures, and equity and efficiency indicators in education sectors were collected from all these secondary sources to analyze equity, efficiency, and other educational trends. Regarding economic analysis, we calculate Lorenz curves and Gini coefficients to measure equity of schooling in both Malawi and Ghana, using the World Bank data set (see <http://www.worldbank.org/research/projects/edattain/>). In addition, recurrent unit costs calculated by other authors<sup>1</sup> are also referred to for comparative efficiency analysis.

---

<sup>1</sup>Calculated by Kunje and Lewin (2000) for Malawi and by Akyeampong, Furlong, and Lewin (2000) for Ghana.

## SITUATION ANALYSIS OF FPE POLICY: MALAWI AND GHANA

Types of fees abolished by FPE policy are different in Malawi and Ghana, as mentioned. The Malawian government commenced FPE policy in 1994 and abolished all fees at once. Fees covered by FPE policy in Malawi are tuition fees, basic textbooks fees, uniform fees, and other direct fees at school (Kadzamira & Rose, 2003). Although indirect fees/costs, especially opportunity cost, are still incurred by parents and children, much of the financial burden was alleviated.

On the other hand, the Ghanaian government initiated FPE policy in 1996. The policy aimed at free and compulsory primary education by 2005 and therefore alleviated school fees/costs gradually. At the initiation of the policy circle, the government embarked on the abolishment of tuition fees officially. Although free-charged tuition fees had been acknowledged in principal by the Education Act and the Constitution before the initiation in Ghana, the policy sought to abolish unsanctioned tuition fees that proliferate at the local level (World Bank, 2004a). Nevertheless, other direct fees/costs such as textbooks, uniforms, stationery, sports kits, and contribution for PTA were still imposed on parents (Avotri, Owusu-Darko, Eghan, & Ocansey, 2000). In 2005, when the policy cycle ended, the rest of other official fees were finally abolished, as Malawi had done since the policy initiation (World Bank, 2006). There are no tuition fees at the public basic level. However private basic schools are not free. Fees in private schools appear to be unregulated, exorbitant, and therefore out of reach of many parents, but in Ghana approximately 13% of total enrolled pupils in 1996 go to private primary schools (World Bank, 2004a).

Educational expenditure as a percentage of government expenditure has fluctuated in both Malawi and Ghana over time (Figure 1). In Malawi, it increased

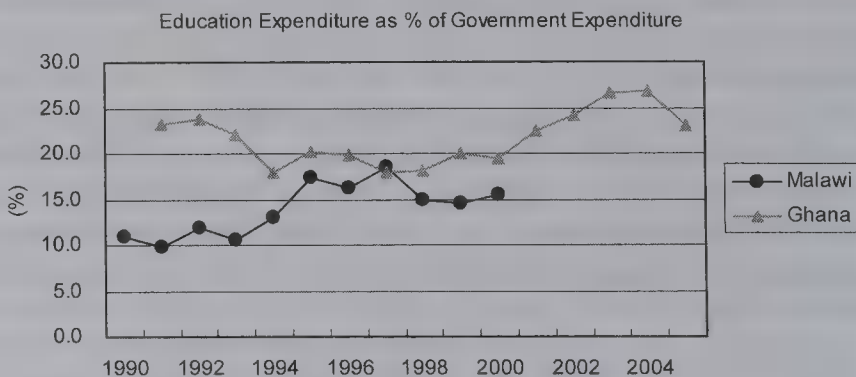


FIGURE 1 Educational expenditure as a percentage of government expenditure in Malawi and Ghana. Source: Ministry of Education in Al-Samarrai (2005) for Malawi; IMF Ghana statistical annex in Foster and Zormeale (2002), and Ministry of Education (2006) for Ghana.

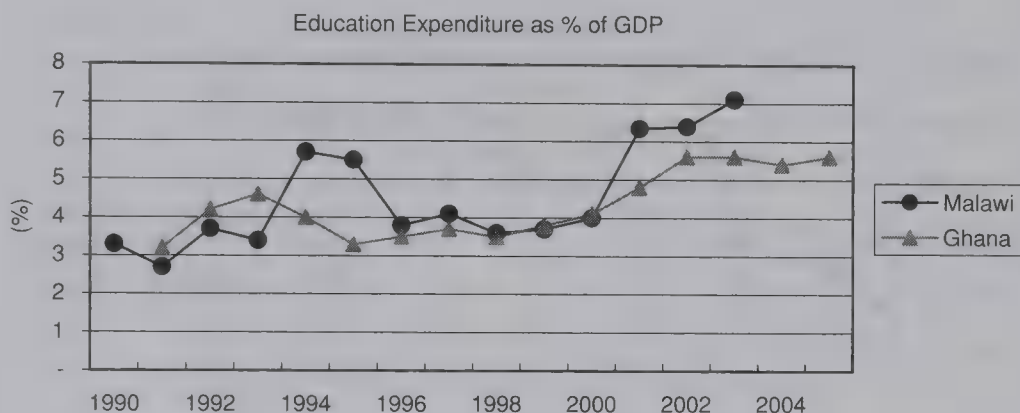


FIGURE 2 Educational expenditure as a percentage of the gross domestic product in Malawi and Ghana. Source: Ministry of Education and Malawi national commission for UNESCO (2004), Ministry of Education (2005) for Malawi; IMF Ghana statistical annex in Foster and Zormele (2002) and Ministry of Education (2006) for Ghana.

gradually from 1994, the year of FPE policy initiation. In Ghana, it reduced around 1996, the initiation of FPE policy, but increased again starting in 2001. Both countries have increased educational expenditure as a percentage of government expenditure, but the Ghanaian government put more emphasis on education financially than did the Malawian government.

The Malawian educational expenditure as a percentage of the gross domestic product (GDP) fluctuates more than its percentage of government expenditure (Figure 2). It increased in 1994, the year of FPE policy initiation; decreased in 1996; and increased again dramatically in 2001. This is because the GDP in Malawi also fluctuated. With regard to Ghana, it steadily increased over time. There are no acknowledged criteria for educational expenditure as a percentage of GDP to judge what percentage is the most profitable investment, but Figure 2 shows that Malawi has invested in education as a ratio of national incomes slightly more than Ghana over time. From the two figures it can be concluded that both countries have increased educational expenditure over time and have a financial commitment to educational development.

### *Change in Access in Primary School (Enrollment)*

FPE policy is expected to raise enrollment, but Malawi and Ghana are not unusual in that. In Malawi, the introduction of FPE in 1994 resulted in an abrupt, massive expansion of enrollments in primary schools (Figure 3). Between 1993 and 1994, the enrollment increased by 51%, from approximately 1.9 million to nearly 2.9 million. This surge of enrollment is the most rapid increase Malawi has



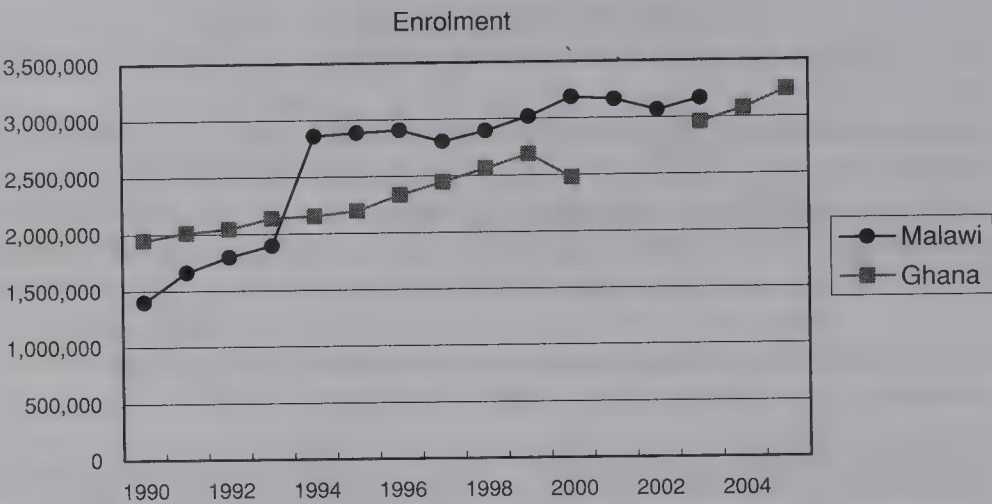


FIGURE 3 Transition of enrollment in Malawi and Ghana. Source: Ministry of Education (2004, 2005) for Malawi; Ministry of Education (2006), and World Bank (2004a) for Ghana.

ever seen. Ghana also increased enrollment but not as dramatically as Malawi. The number of pupils has been rising gradually.

EQUITY ANALYSIS

The Impact on Equity in Schooling

*Access to schooling.* This section explores how the change in policy and financing for FPE actually made an impact on equity in enrollment, analyzing the change in gross enrollment rate (GER) in primary education. According to Table 1, in Malawi the access to primary education drastically expanded after 1994, and the GER in all quintiles increased. The increase in all quintiles reached around 120%

TABLE 1  
Gross Enrolment Rates by Household Income Quintile in Malawi and Ghana

	Poorest 20%	2nd	3rd	4th	Richest 20%	Total	
Malawi	1990	58	76	86	97	110	81
	1997	117	121	119	125	120	120
Ghana	1992	75	91	90	91	101	88
	1997	70	83	85	90	94	84

Source: Al-Samarrai and Zaman (2002) and Sudharshan and Xiao (2002).

TABLE 2  
Gross Enrolment Rates by Gender and Region in Malawi  
and Ghana

		<i>Male</i>	<i>Female</i>	<i>Urban</i>	<i>Rural</i>	<i>Capital</i>
Malawi	1990	86	75	115	77	/
	1997	128	113	119	120	/
Ghana	1992	93	83	97	84	99
	1997	87	80	92	80	95

Source: Al-Samarrai and Zaman (2002) and Sudharshan and Xiao (2002).

and eliminated a large disparity in GER. On the other hand, the GER in Ghana did not increase or even change in all income quintiles or the total. Accordingly there was no change in disparity in GER in Ghana. On the contrary, for the poorest 20% the GER decreased in 1997. The disparities in GER by gender in the two countries were not alleviated, even though the GER in each quintile increased on average (Table 2). Whereas the gap in GER by region did not change in Ghana, the GER in rural areas in Malawi reached up to the standard of that in urban areas and removed the disparity between rural and urban areas.

*Attainment in schooling.* This section explores equity in the school attainment in Malawi and Ghana, examining Gini coefficient. The attainment rates used for calculating the Gini coefficient in this section is from Grade 1 to Grade 9. Malawi has 8 years in primary education and 2 years in secondary education, and Ghana has 6 years in primary education and 3 years in junior secondary education. Therefore, Grade 9 corresponds to the 1st year in secondary education in Malawi and the 3rd year in junior secondary education in Ghana. This section does not measure the attainment in each education level such as primary or secondary level because the length of years in each education level in Malawi and Ghana is different and it is not fair to compare them. However, that enables us to examine equity overall, beyond the level of school. Tables 3 and 4 show the results of the calculations for Gini coefficients. The tables indicate that Ghana achieves better in school attainment than Malawi as a whole.<sup>2</sup> The Gini coefficient of the “total” in 2003 in Ghana is 0.25, whereas in 2000 in Malawi it is 0.3. Equity, however, is investigated by focusing on the disparity among cohorts, as analyzed for enrollment earlier. Malawi improved attainment drastically and alleviated disparity, although some disparity remained (Table 3). In contrast, Ghana did not improve attainment from 1993 to 2003 and reduced disparity among cohorts very little

<sup>2</sup>The smaller the Gini coefficient is, the “better” it is for attainment.

TABLE 3  
Malawi: Gini Coefficient for School Attainment by Income Quintile, Gender and Region

Year	Total	Quintile			Gender		Region	
		Poorest 40%	Middle 40%	Richest 20%	Male	Female	Urban	Rural
1992	0.46	0.53	0.47	0.30	0.41	0.51	0.28	0.48
2000	0.30	0.31	0.31	0.20	0.29	0.30	0.17	0.31

Note: Calculated from the World Bank data set (2006).

Source: EdStats obtained at <http://www.worldbank.org/research/projects/edattain/>.

(Table 4). Focusing on the poorest quintile, the attainment in that quintile became worse and the disparity among income quintiles clearly widened in 2003. In both countries, disparity within regions still remained with the same proportions.

### Discussion on Equity in Schooling

*Analysis of the impact on equity in schooling.* The change in policy and financing for FPE made different impacts on equity in schooling in Malawi and Ghana. Malawi reduced disparities in both enrollment and attainment among income quintiles and enhanced equity. The disparity between regions in Malawi was eliminated in enrollment level but not in attainment level. The disparity in gender in Malawi was not reduced in either enrollment or attainment level. On the other hand, Ghana had worse results in terms of equity. In enrollment and attainment in all categories, the disparities were not lessened. There was no positive impact on equity in Ghana.

To analyze the impacts on equity in schooling aggregately, Malawi, which focused on quantitative expansion in FPE policy, increased access totally but especially the access of disadvantaged cohorts and reduced the gap in enrollment and attainment. Although Ghana, which put an emphasis more on qualitative expansion at first in FEP policy, maintained the higher standard of quality, which

TABLE 4  
Ghana: Gini Coefficient for School Attainment by Income Quintile, Gender and Region

Year	Total	Quintile			Gender		Region	
		Poorest 40%	Middle 40%	Richest 20%	Male	Female	Urban	Rural
1993	0.28	0.32	0.31	0.14	0.25	0.30	0.18	0.34
2003	0.25	0.35	0.22	0.13	0.23	0.27	0.17	0.32

Note: Calculated from the World Bank data set (2006).

Source: EdStats obtained at <http://www.worldbank.org/research/projects/edattain/>.



TABLE 4.1  
Gross Enrolment Rate in Malawi and  
Ghana

	Malawi	Ghana
1990	75	79.3
1991	79	79
1992	85	77.6
1993	91	78.1
1994	127	75.9
1995	115	74.6
1996	132	76.5
1997	128	77.5
1998	128	78.4
1999	131	79.4
2000	111	78.6
2001	/	80
2002	/	/
2003	/	86.5
2004	/	87.5
2005	/	92.1

Source: MoE in Al-Samarrai 2005 for Malawi, MoE (2006) and World Bank (2004a).

was indicated by higher level of attainment, it could not reduce disparity even in attainment level. Enrollment expresses registration and is increased by quantitative expansion. Attainment expresses achieved rate and is maintained by quality of education. The cases of the two countries indicate that quantitative expansion alleviated disparity and enhanced equity in schooling, but quality improvement did not contribute to reduce the disparities within each category.

*From equity in resource allocation to equity in school.* FPE policy is expected to improve equity in the allocation of public resources and the opportunity for schooling, but how and in what points did Malawi and Ghana enhance equity by FPE policy? Malawi changed the allocation of resources favorably for the poorer households, providing a larger proportion of subsidies to them than the average. This pro-poor allocation is deemed to be equitable allocation, based on the concept of procedural equity; if children from low-income households can be assumed to have more financial constraints for schooling, they need some “positive discrimination” for more equal opportunity. Moreover, this equitable allocation of subsidies contributed toward achieving nearly equal distribution in school enrollment and attainment among income quintiles. The concept of distribution equity—“equals” should be treated “equally”—was achieved among income quintiles in Malawi.

On the other hand, Ghana eliminated the disparity in resource allocation but did not support the poorer quintiles more than other, richer quintiles. Equality in resource allocation was achieved, but there was no impact on equity in enrollment and attainment. Some field researches report that 48.1% of children has monetary cost as the biggest reason for not attending school in Ghana (World Bank, 2004a), whereas 24.1% in Malawi had that reason (World Bank, 2004b). Considering this situation, more financial support for the poorer households such as eliminating other school costs or scholarship should be given in Ghana as well as in Malawi to improve equity in schooling. One of the reasons for the disparity, especially in attainment in Ghana, is the difference in quality of education between public and private school. The difference of quality affects the difference of survival and achieved rate. The promotion of privatization for cost recovery is important. However, the huge gap of quality between public and private school is problematic for equity because only financially richer children have the opportunity to go to a private school, which offers better quality education. The budget, which is saved by the cost recovered from the promotion of privatization, should be spent more on the poorer quintiles if it is to have an actual impact on equity in schooling.

In addition to household income quintiles, the introduction of FPE policy changed the allocation of public resources more equitably to a certain degree among gender and geographical regions in Malawi and Ghana. However, equity did not take root in schooling. Cost reduction, especially with pro-poor policy, affected the expansion of access, as seen in Malawi, but did not make a critical contribution toward alleviating disparity in schooling in either country.

## EFFICIENCY ANALYSIS

The FPE program covers school costs for all by increasing expenditure, and it encourages all children to participate in primary schools. Because the FPE program spends more expenditure on “all” children, it inevitably contributes to improving equity. At the same time, because it subsidises all children, governments need to allocate more of their budget to the education sector. In this sense, the introduction of FPE policy is a big challenge for national governments, and they cannot spend the budget wastefully. But how efficiently do national governments use resources in introducing FPE policy? Does FPE policy have some impact on efficiency in schools? This section analyzes how efficiently resources are allocated over time and the impact of FPE policy on efficiency in resource use in Malawi and Ghana.

### *Efficiency in Resource Allocation*

*Allocation of total recurrent education expenditure.* This section explores how recurrent education expenditure is allocated in Malawi and Ghana. Figures 4 and 5 show the allocation of recurrent expenditure to each level of education in Malawi and Ghana, respectively. From the figures, it can be seen that the

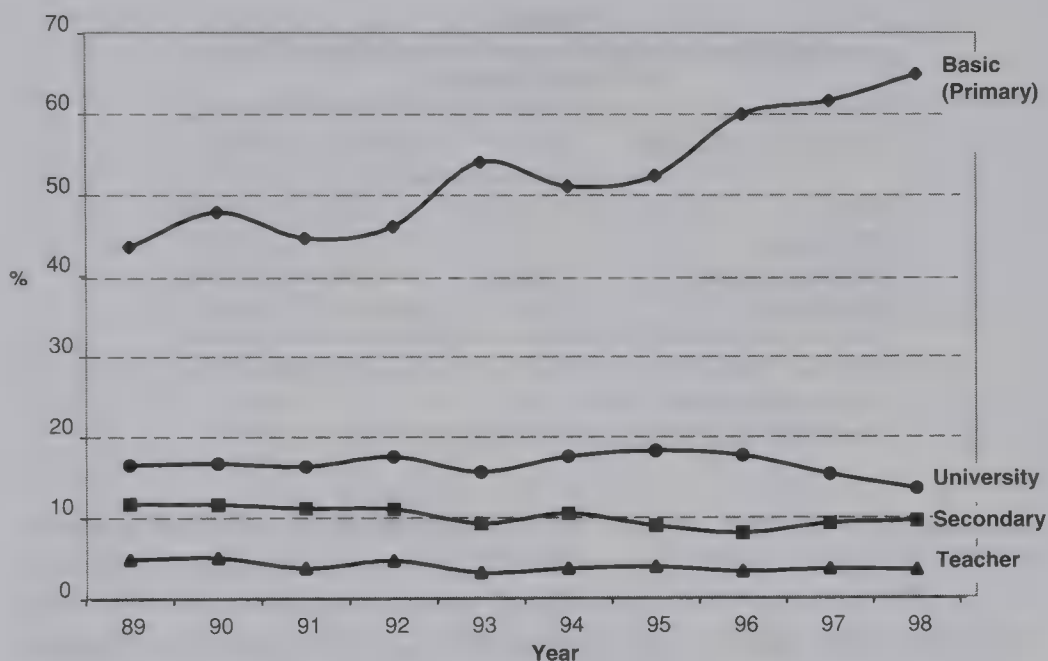


FIGURE 4 Malawi: Allocation of recurrent public expenditure by level of education—percentage of total. As cited in Kunje, Lewin, and Stuart (2002, p. 4).

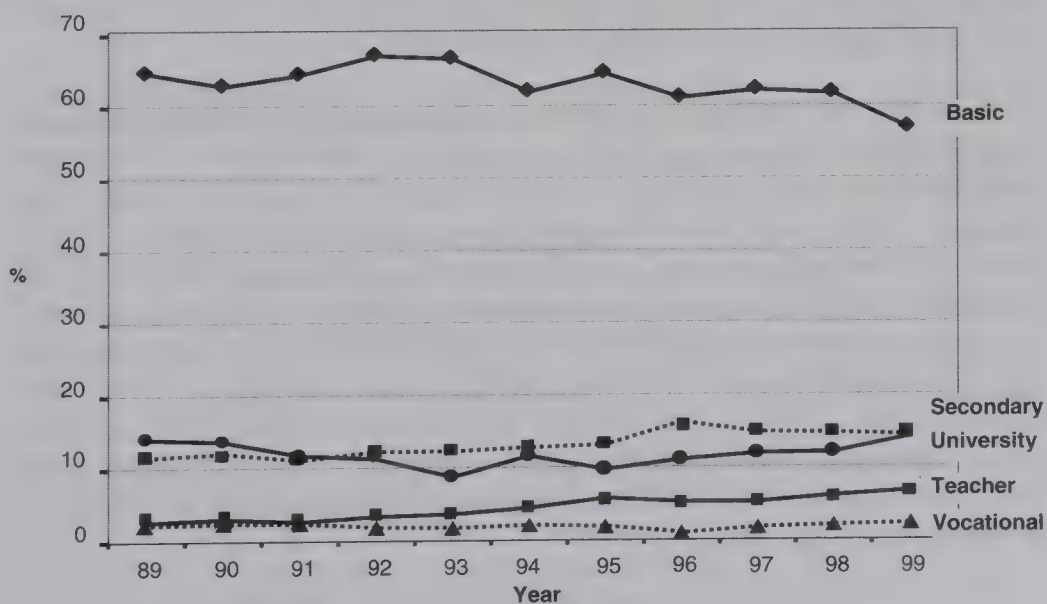


FIGURE 5 Ghana: Allocation of recurrent public expenditure by level of education—percentage of total. As cited in Akyeampong, Furlong, and Lewin (2000, p. 4).

TABLE 5  
Malawi: Recurrent Public Expenditure Per Student (Constant  
US\$ in 1991)

<i>Level/Type of Education</i>	<i>1993</i>	<i>1994</i>	<i>1996</i>
Primary	19.50	18.38	16.88
Secondary	163.50	153.75	115.13
Teacher education	584.63	692.25	919.50
University	3090.75	3265.50	3744.00

*Note:* Adjusted for inflation using constant 1991 US\$ prices (US\$1 = MK2.664).

Source: Calculated from data of MOE by Kunje & Lewin (2000).

distinctions between them are threefold. First, Malawi increased the proportion of recurrent expenditure for basic level after FPE policy, and it reached approximately 65% in 1998, whereas Ghana had reduced that for basic level after FPE policy and reached approximately 57%. Both countries, however, keep the proportion for basic level at more than 50%. Second, in Malawi the allocation to university is higher than to secondary education, but by contrast, in Ghana the allocation to secondary education has been higher than to university. This is because the proportion of allocation to secondary education is relatively low in Malawi and the proportion saved is allocated to basic education. Third, the proportion for teacher education has stagnated in Malawi, but it has increased gradually in Ghana. This indicates the degree of advance preparation for FPE policy. There is a need to hire many more teachers with a sudden increase in pupils.

*Recurrent expenditure per student.* The proportion of recurrent allocation to each level of education has been examined in an earlier section. This section investigates how much each level of education spends in recurrent expenditure on a student. Tables 5 and 6 show recurrent public expenditure per student—what is called recurrent unit cost—by level of education, a point which should be paid attention to for efficiency analysis is the ratio of primary education unit cost to each other education unit cost. Table 7 calculated from Tables 5 and 6 shows the ratio: how many times unit cost of each level is greater than that in primary level in each country. As seen from Table 7, the unit costs of other levels are much higher than that of primary level in Malawi, compared to those in Ghana. In this case, the wide gap indicates that the government spends too much on other levels of education per person.

### *Impact on Efficiency in Resource Use*

This section investigates how policy and financing for FPE made an impact on efficiency in resource use in basic education. To investigate that, this section



TABLE 6  
Ghana: Recurrent Public Expenditure Per Student  
(Constant US\$ in 1996)

<i>Level/Type of Education</i>	<i>1992</i>	<i>1995</i>	<i>1998</i>
Primary	36.79	44.25	41.75
JSS	66.76	86.55	67.96
SSS	77.44	153.88	168
Vocational/Technical	188.37	139.04	299.54
Teacher education	246.62	442.6	617.31
University	1376.94	1123.87	855.91

*Note:* Adjusted for inflation using constant 1996 US\$ prices (US\$1 = c1637).

Source: Calculated from data of MoE and World Bank by Akyeampong, Furlong & Lewin (2000).

explores how human resources (teachers) and learning materials (textbooks and classrooms) are organized. In addition, to analyze how efficiently learning is provided without losing students who drop out (called *wastage*), the next section examines changes in repetition rate, survival rate, completion rate, and transition rate over time.

As discussed earlier, primary education in Malawi has 8 years as compulsory basic education. Primary and junior secondary education in Ghana has 6 years and 3 years, respectively. The final grade in junior secondary school (JSS) in Ghana is still Grade 9, and therefore JSS is still regarded as compulsory “basic education.” To compare the student flow in the two countries fairly, primary education and JSS are analysed as basic education for the case of Ghana.

TABLE 7  
Ratio of Primary Level's Unit Cost and Other Level's  
Unit Cost in Malawi and Ghana

<i>Malawi</i>	<i>1996</i>	<i>Ghana</i>	<i>1998</i>
Primary	1	Primary	1
		JSS	2
Secondary	7	SSS	4
		Vocational/ technical	7
Teacher education	54	Teacher education	15
University	222	University	21

*Note:* Calculated from Tables 5 and 6. The figures represent the ratio of primary level's unit cost and other level's unit cost.

*Teachers and trained teachers.* Human resources are the important factors that are directly related to providing pupils with quality learning. In particular, whereas FPE policy is expected to encourage more children to go to primary school, the supply of more teachers to primary school is also critical to maintain quality learning. Thus, when FPE policy is implemented, teacher supply also should be well planned to meet “demand” of increased pupils efficiently. This section explores efficiency in the relationship between supply and demand of teachers along with the introduction of FPE.

Figure 6 shows the changes in the number of pupils and teachers over time in primary education in Malawi and Ghana. As seen in the figure, Malawi increased the number of enrollments in primary school in 1994 with the introduction of FPE. The total number of pupils surged from 1.9 million to 2.9 million. It rose to about 1.5 times more than the previous year, 1993–1994. As the number of pupils increased in school in 1994–95, the Malawian government also provided schools

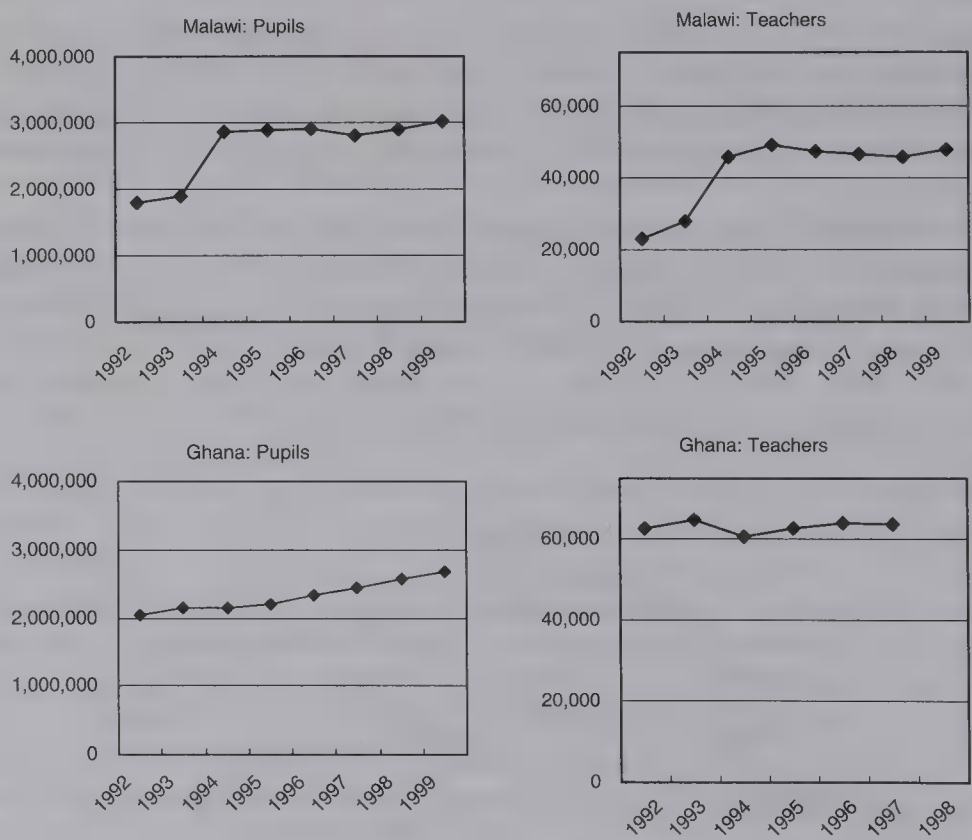
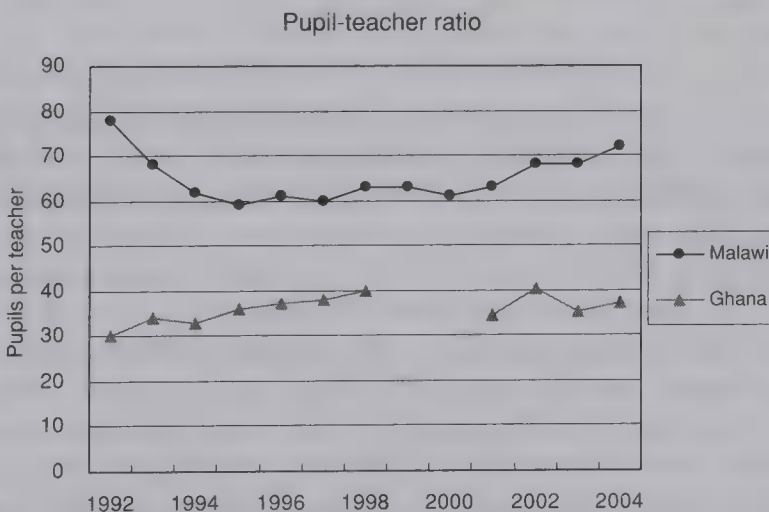


FIGURE 6 Transition of number of pupils and teachers in primary education in Malawi and Ghana. Source: Ministry of Education (2004, 2005) for Malawi; Ministry of Education in Akyeampong and Furlong (2000), Ministry of Education (2006), and World Bank (2004a) for Ghana.

with teachers promptly. The number of teachers grew from 28,000 to 46,000 and increased 1.6 times as much as the last year, as seen in the figure. In contrast, the number of pupils enrolled in primary school in Ghana increased gradually over time. The supply of teachers in Ghana has been nearly stable over time. With regard to the difference in absolute number of pupils and teachers, the number of pupils in Malawi increased drastically after 1994 and exceeded that in Ghana. However, the number of teachers in Malawi has stayed lower than Ghana over time: The number of teachers in Ghana is still more than 60,000. This means that even though the number of teachers surged in Malawi, there are still fewer teachers relative to Ghana.

To analyze the balance of the number of pupils and teachers more closely, pupil–teacher ratio is referred to (see Figure 7). The figure indicates that the number of teachers increased over time around the commencement of FPE policy as previously discussed, and pupil–teacher ratio became lower. Hence, it can be seen that the learning environment, which is encouraged by the interaction between pupils and teachers, was more efficiently managed than before the commencement of FPE policy. In Ghana, the number of pupils did not surge abruptly and teachers were provided constantly; therefore the pupil–teacher ratio has also been stable—between 30 and 40. In addition to the fluctuation over time, the figure indicates the quality of the learning environment. It indicates the poorer learning environment over time with a high pupil–teacher ratio in Malawi compared with Ghana.



**FIGURE 7** Transition of pupil–teacher ratio in primary education in Malawi and Ghana.  
*Note:* The data for 1999 and 2000 for Ghana are not available. Source: Ministry of Education (2005), World Bank (2004b) for Malawi; Ministry of Education in Sudharshan and Xiao (2001) and Ministry of Education (2006) for Ghana.

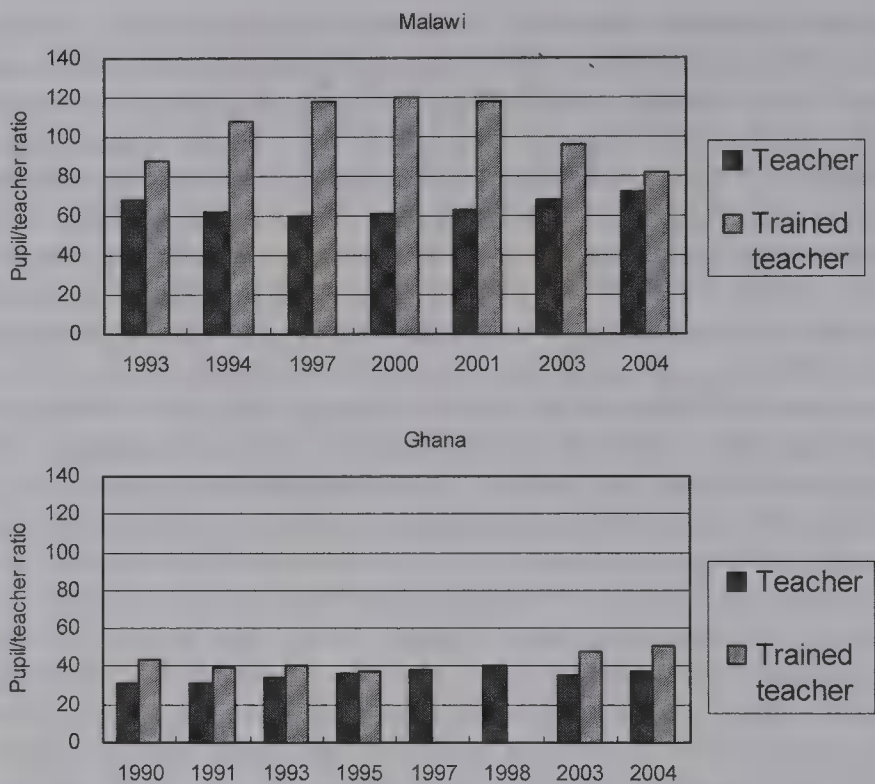


FIGURE 8 Pupil–teacher and pupil-trained teacher ratios in primary education in Malawi and Ghana. Source: Ministry of Education (2005), World Bank (2004b) for Malawi; Ministry of Education in Sudharshan and Xiao (2001) and Ministry of Education (2006) for Ghana.

Figure 7 pupil–teacher ratios show the balance in the number of pupils and teachers. Next, Figure 8 clarifies pupils-trained teacher ratios as well as pupil-teacher ratios in Malawi and Ghana. With regard to Malawi, pupil-trained teacher ratio is extremely high, whereas pupil–teacher ratio is stable. After 1994, the ratio reached more than 110:1 in 1997, 2000, and 2001. These numbers indicates that although teachers were supplied to schools after the introduction of FPE and kept the pupil–teacher ratio stable, the teachers supplied urgently, as pupil numbers increased, were not trained. In 1997, 2000, and 2001, nearly half of teachers were untrained. In Ghana, on the other hand, pupil–teacher and pupil-trained teacher ratios are stable and nearly the same figures. This means that, first, Ghana supplied a high proportion of trained teachers originally, even before the FPE program began, and second, the ratios are not affected badly by the commencement of FPE policy. In terms of the demand and supply of teachers, especially high-quality teachers, human resources are managed more efficiently for a better learning environment in Ghana than in Malawi.



TABLE 8  
Pupil-Textbook (Basic Subjects) Ratio  
in Malawi and Ghana

	<i>Malawi</i>	<i>Ghana</i>
1993	3	/
1994	7	/
1997	3	/
2002	/	2.0
2003	/	1.3
2004	1.5	1.0

*Note:* “/” means “not available” (n.a.)

Source: MoE in Kadzamira and Rose (2003) and MoE (2005) for Malawi, MOE in the World Bank (2004a) for Ghana.

*Education materials.* Even if more pupils come to participate in primary school, given lack of learning materials, the learning is not efficiently carried on. Along with the expectation of increase of pupils, a sufficient supply of learning materials and environment is necessary for the states. In Malawi, the abrupt surge of pupils led to a shortage of learning materials after FPE programs began. Table 8 indicates the critical shortage of textbooks after 1994. In 1994, only one textbook was available to about 7 pupils in primary school. In the year following the commencement of FPE policy, the availability rate of textbooks became less than half relative to the previous year. With regard to permanent classroom, which means a building and does not include only space under a tree outside, Malawi ran short of enough classrooms after 1994.

Table 9 shows the ratio of 162 pupils per permanent classroom. This is a distinct lack of resources for quality learning. There are no data on textbooks and

TABLE 9  
Pupil-Permanent Classroom Ratio in Malawi  
and Ghana

	<i>Malawi</i>	<i>Ghana</i>
1993	102	/
1994	162	/
1997	156	/

*Note:* “/” means “not” available (n.a.).

Source: MoE in Kadzamira and Rose (2003).

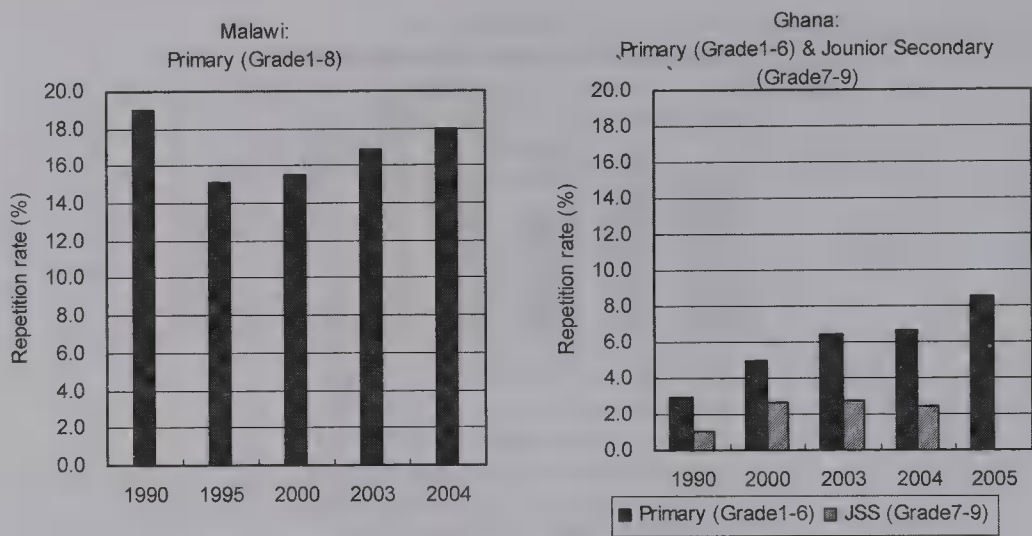


FIGURE 9 Repetition rate in basic education in Malawi and Ghana. Source: Ministry of Education (2005) for Malawi; Ministry of Education (2006) and World Bank (2004c) dataset for Ghana.

classrooms before the commencement of FPE available for Ghana, but as Table 8 indicates, Ghana kept the pupil–textbook ratio at a lower rate than Malawi.

***Repetition rate.*** In this section student flow in school is explored to analyze how efficiently learning is provided without wastage.

Figure 9 shows repetition rates in basic education in Malawi and Ghana. As the figure indicates, the repetition rate of Malawi is much higher over time than that of Ghana. Therefore, in terms of educating more pupils smoothly with limited resources and costs, Malawi, which has 15% of people studying in the same grade again, can be regarded as inefficient. However, regarding the change in repetition rate before and after FPE, the repetition rate in Malawi decreased a little after that, although it has increased again gradually. Ghana has an increased repetition rate in primary school, maintaining a low repetition rate in higher level in basic education, JSS. Ghana began to abolish more school costs such as uniform and books in basic education in 2005, following the elimination of tuition fees in 1996. The reason for the additional increase in 2005 was considered to be because the abrupt increase of pupils made for a higher proportion of repeaters. In contrast, although Malawi lessened the repetition rate in 1995, following the year of the introduction of FPE, it is thought that a certain part of the constraints that encourage repetition was school costs. This was alleviated by the FPE program.

FPE policy affects the balance among components consisting of schooling such as quantity, quality, and costs, as discussed earlier, and the change in balance makes efficiency of schooling higher or lower.

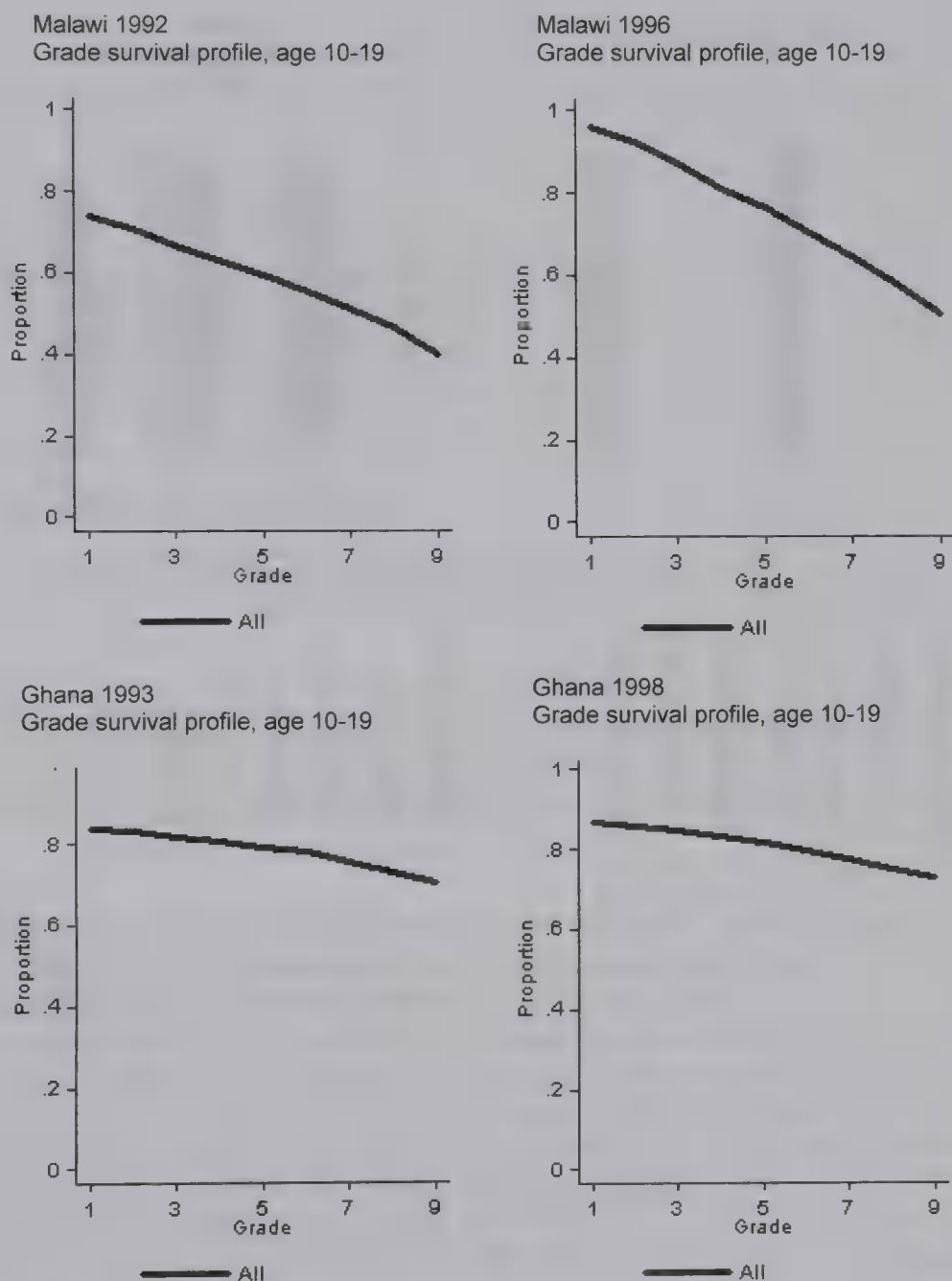


FIGURE 10 Grade survival profiles in Malawi and Ghana. Source: World Bank (2006), analyzing the data from DHS.

**Survival rate.** As Figure 10 shows, the trends of the change in grade survival profiles over time in Malawi and Ghana are different. Malawi improved the survival rate aggregately in each grade. The biggest improvement lies in the survival rate in the first grade, and more pupils survive and transit to the next grade. In Ghana,

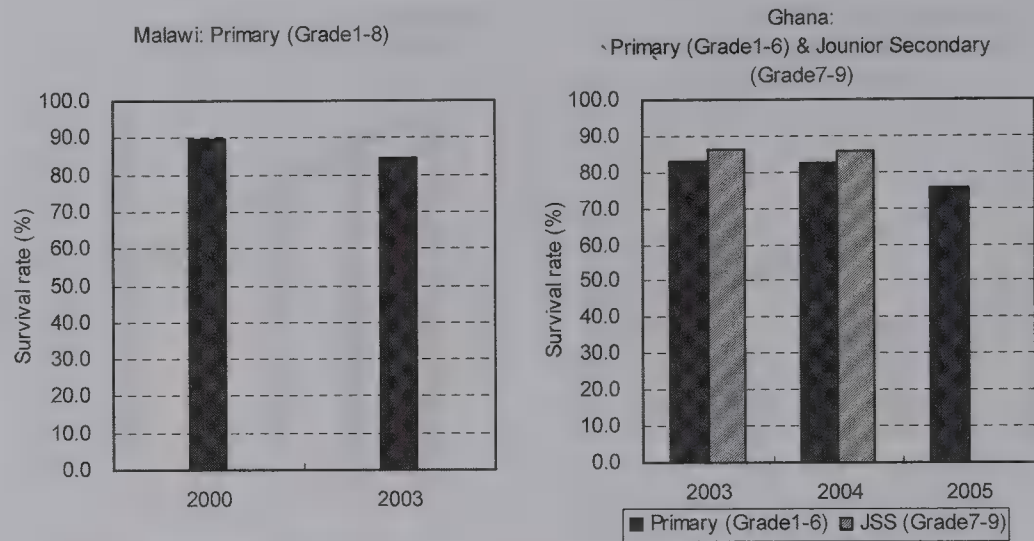


FIGURE 11 Survival rate in basic education in Malawi and Ghana. Source: Ministry of Education (2005) for Malawi; Ministry of Education (2006) for Ghana.

the survival rate was not improved at all after the commencement of FPE policy. The slopes of survival rates after FPE policy are different in Malawi and Ghana, but the averages of survival rates are quite similar between Malawi and Ghana (Figure 11).

**Completion rate.** The standard of completion rate in Malawi and Ghana is different (Figure 12). Malawi has had a low completion rate over time. Malawi increased the completion rate gradually and did not show a drastic impact on it, at least on a short-term basis, after the introduction of FPE. It did gradually increase but is still low. The completion rate was about 25% in 1992 and about 40% in 2000 in Malawi. The completion rate in Ghana has been higher than in Malawi but still not sufficient, considering that basic education is regarded as “compulsory.” The completion rate in Ghana also increased gradually, but it rose higher in 2005. This is also seen as the influence of the reduction of financial constraints by FPE policy. This increase had a sudden impact in both countries.

**Transition rate.** Transition rate also indicates the degree of internal efficiency for the whole education sector. If the transition rate is low, it indicates the connection from primary to secondary level is not well structured. This education structure without appropriate provision redressing the balance of the



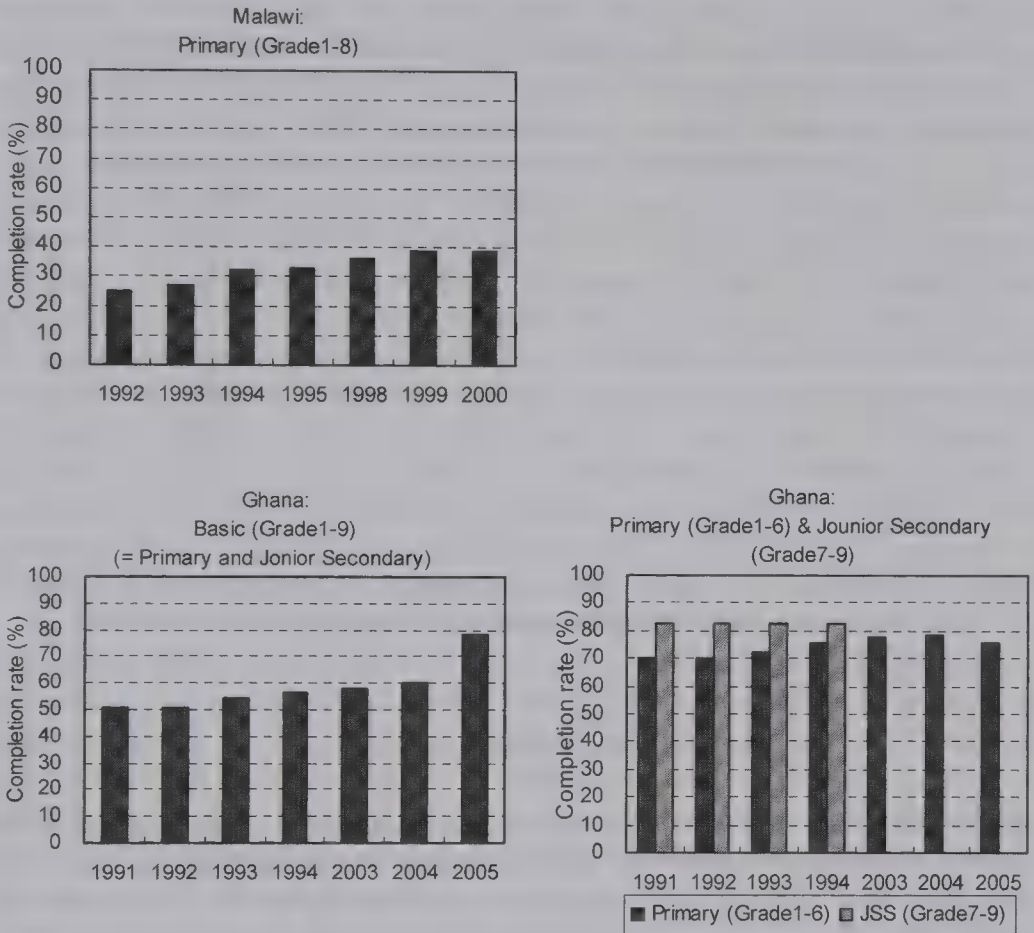


FIGURE 12 Completion rate in basic education in Malawi and Ghana. Source: World Bank (2004b) for Malawi; Ministry of Education (2006) for Ghana.

number of pupils between output from primary and input to secondary level generates some wastage of students who miss out on possibilities of enhancing themselves at the next level of education. Table 10 shows the transition rate from primary to secondary level of education in Malawi and Ghana. Although the transition rates before the introduction of FPE in Malawi are not available in the table, Malawi maintained 75% in 2000 and 51% in 2003. On the other hand, the transition rate from JSS to Senior Secondary School (SSS) in Ghana, actually from basic level to secondary level, is from 33% to 47% and indicates a lower rate than Malawi. However, Ghana has established high transition rates over time. Ghana increased transition rate a little more after the commencement of FPE.

TABLE 10  
Transition Rate from Primary Level to Secondary Level in Malawi and Ghana

	<i>Malawi</i>	<i>Ghana</i>	
	<i>Primary to Secondary (Grade 8 to 9)</i>	<i>Primary to Junior Sec (Grade 6 to 7)</i>	<i>Junior Sec to Senior Sec (Grade 9 to 10)</i>
1991	/	96.8	/
1992	/	93.9	/
1993	/	95.0	/
1994	/	94.5	/
2000	75.0	96.0	33.2
2001	/	/	41.5
2002	/	/	47.3
2003	51.0	/	/
2004	/	97.1	/

*Note:* means “not available” (n.a.).  
Source: MoE (2005) for Malawi, Government of Ghana (2003), MoE (2006) and World Bank (2004c) for Ghana.

CONCLUSION

Malawi and Ghana introduced FPE policy in 1994 and 1996, respectively, at early stages in the context of EFA. The two countries, however, had different emphases regarding “free” primary education in their policies. In the case of Malawi all school fees such as tuition, uniform, and textbook fees were abolished by FPE policy. On the other hand, Ghana abolished a part of school fees—only tuition fees—in the commencement of FPE policy and had a preparation period for a gradual reduction of other school fees. Consequently, the difference in the degree of the reduction of school fees, namely, the degree of government subsidies, led to different trends income quintiles in distribution of resources and its impacts on schooling in terms of equity and efficiency.

In Malawi, as a result of the greater degree of subsidies for school fees, the disparities among income quintiles in distribution of resources were removed and the situation became more pro-poor and equitable. Moreover, the equitable distribution with increased access, especially of the poor quintile, made a greater impact on reducing disparity rapidly and enhancing equity satisfactorily in enrollment and attainment measures. This substantial improvement on equity in resource allocation and schooling is the highly advantageous outcome of FPE policy in Malawi. However, the trade-off of the drastic change for equitable subsidisation lay in inefficient resource allocation and use. The sudden abolishment of all formal school fees led to a rapid increase in enrollment but at the same time created a

huge demand for teachers. Malawian government urgently supplied unqualified and untrained teachers, and most of the large budget increase was spent on their wages rather than on other necessary provisions. For example, the heavy spending on untrained teachers' wages from increased recurrent expenditure led to a reduction of capital expenditure and then to a shortage of supply of learning resources such as textbooks and classrooms. In addition, the small proportion of budget on teacher training over time caused the undersupply of trained teachers. In Malawi, the proportion of recurrent expenditure to teacher training has not increased in spite of increased demand. After the introduction of FPE policy, in 1994, pupils-trained teacher ratio, pupil-textbook ratio, and pupil-permanent classroom ratio were 108, 7.1, and 162, respectively. Thus, the FPE program in Malawi lost the balance between recurrent and capital expenditure and accordingly between supply and demand in trained teachers and learning materials. Furthermore, public expenditure on other higher levels of education such as secondary and tertiary education is still relatively too large, and Malawi needs to promote more cost reduction and cost sharing for efficient resource allocation in the education sector.

In contrast, Ghana, which abolished only tuition fees in the commencement of FPE policy, did not have great "confusion" about resource allocation, and technical efficiency is likely to be maintained. Ghana has been increasing the proportion of the budget to teacher training gradually in advance and managed to supply sufficient trained teachers continuously. Also, no specific inefficiency appears in budgetary allocation to each level of education, at least relative to Malawi. However, FPE policy in Ghana did not contribute to making a big improvement for equitable distribution of resources and equity in schooling. For example, among income quintiles, most disparities were removed and equality was secured in distribution of resources, but disparities in schooling were not sufficiently alleviated. On the contrary, GER and attainment measured by Gini coefficient deteriorated in the poorest quintile even after the commencement of FPE policy and gaps widened slightly among other richer quintiles. Considering this situation that "equal" distribution did not make a sufficient impact on actual equity in schooling, subsidising poor households more with "positive discrimination" may be required for the achievement of "equity" in schooling, based on the concept of procedural equity and on the rationales by EFA and MDGs. It has not yet been clarified whether more subsidies to abolish other school fees in addition to tuition fees, or student loan, or other types of grants are the best means to enhance both equity and efficiency aggregately, because this is not the objective of this research. Nevertheless this research discovered that the case of Malawi enhanced equity by covering more fees. Moreover if some points were attended to, it would have been possible to increase efficiency as well as equity.

The following are implications for financing FPE policy learned from the experiences of Malawi and Ghana.

- If governments emphasize reducing inequity further in schooling, the abolition of tuition fees and other fees for poor households is one effective way to enhance equity. Abolition of only tuition fees by FPE policy removed the disparities within income quintiles in resource allocation but did not remove them in actual schooling.
- Other provisions as well as FPE policy are required to reduce disparities by gender and geographical region in schooling. Neither country made big or sufficient changes in equity through FPE policy.
- For efficient allocation and use of resources, governments should allocate sufficient budget for teacher training in advance of the commencement of FPE policy and prepare to supply more teachers for future increases in pupil numbers.
- For efficient use of public resources, governments should promote cost sharing, cost reduction, and privatization to a certain level, especially at higher levels of education. Otherwise, financial constraints will remain and the FPE program would not be sustainable if the subsidies from international donors are cut or abolished.
- To avoid the confusion in financing, especially against the abrupt loss of the balance between demand and supply, the sudden introduction of FPE policy eliminating all fees should be planned sufficiently in advance or even avoided.
- Governments should ensure good provision for quality of education as well as FPE programs. Efficiency indicated by student flow in education system refers to quality of education, but the student flow has not been improved much in the two countries.

## REFERENCES

- Akyeampong, K., & Furlong, D. (2000). *Ghana: a baseline study of the teacher education system*. Brighton, UK: University of Sussex, Institute of Education, Centre for International Education.
- Akyeampong, K., Furlong, D., & Lewin, K. (2000). *Costs and financing of teacher education in Ghana*. Brighton, UK: University of Sussex, Institute of Education, Centre for International Education.
- Al-Samarrai, S. (2005). *Financing primary education for all: public expenditure and education outcomes in Africa*. London: Department for International Development.
- Al-Samarrai, S., & Zaman, H. (2002). *The changing distribution of public expenditure in Malawi* (World Bank Africa Region Working Paper, Vol. 29). Washington, DC: World Bank.
- Avotri, R., Owusu-Darko, L., Eghan, H., & Ocansey, S. (2000). *Gender and primary schooling in Ghana*. Brighton, UK: University of Sussex, Institute of Development Studies.
- Bentaouet-Kattan, R., & Burnett, N. (2004). *User fees in primary education*. Washington, DC: World Bank, Human Development Network, Education Sector.
- Birdsall, N. (1982). *Child schooling and the measurement of living standards*. Washington, SC: World Bank.
- Colclough, C., & Al-Samarrai, S. (1998). *Achieving schooling for all: budgetary expenditures on education in sub-Saharan Africa and South Asia*. Brighton, UK: University of Sussex, Institute of Development Studies.



- Foster, M., & Zormeale, D. (2002). *How, When and does poverty get budget priority? Poverty reduction strategy and public expenditure in Ghana*. London: Overseas Development Institute.
- Government of Ghana. (2003). *Ghana poverty reduction strategy 2003–2005, An agenda for growth and prosperity (Vol. 1 Analysis and policy statement)*. Ghana, Accra: Author.
- Johnes, G. (1993). *The economics of education*. Houndmills: Macmillan.
- Kadzamira, E. C., & Rose, P. (2003). Can free primary education meet the needs of the poor?: Evidence from Malawi. *International Journal of Educational Development*, 23, 501–516.
- Kunje, D., & Lewin, K. (2000). *The costs and financing of teacher education in Malawi*. Brighton, UK: University of Sussex, Institute of Education, Centre for International Education.
- Kunje, D., Lewin, K., & Stuart, J. (2002). *Primary teacher education in Malawi: Insights into practice and policy*. Brighton, UK: University of Sussex, Institute of Education, Centre for International Education.
- Le Grand, J. (1991). *Equity and choice: An essay in economics and applied philosophy*. London: Routledge.
- Levacic, R. (2005). *Economic value of education: Notes on equity*. London: Institute of Education, University of London.
- Mace, J. D. (1996). *Economics of education*. London: University of London.
- Mace, J. D. (2000). *Economics of education* (2nd ed.). London: University of London.
- McMahon, W. W. (1982). Efficiency and equity criteria for educational budgeting and finance. In W. McMahon & T. G. Geske (Eds.), *Financing education: Overcoming inefficiency and inequity*. (pp. 1–30) Urbana: University of Illinois Press.
- McMeekin, R. W. (1975). *Educational planning and expenditure decisions in developing countries, with a Malaysian case study*. New York: Praeger.
- Mingat, A., & Tan, J. P. (1985). *Improving the quantity-quality mix in education: A simulation of policy tradeoffs* (World Bank Education and Training Series Discussion Paper No. 15). Washington, DC: World Bank.
- Ministry of Education (Ghana). (2006). *Preliminary education sector performance report 2006*. Ghana, Accra: Ministry of Education, Science and Sports Ghana.
- Ministry of Education (Malawi). (2005). *Update of the education chapter of the Public Expenditure Review 2000 final draft*. Malawi, Lilongwe: Author.
- Ministry of Education (Malawi) and Malawi National Commission for UNESCO. (2004). *The development of education in Malawi 2004 report*. Ministry of Education and the Malawi National Commission for UNESCO, Lilongwe.
- Monk, D. H. (1990). *Educational finance: An economic approach*. New York: McGraw-Hill.
- Musgrave, R. A. (1959). *The theory of public finance: a study in public economy*. New York: McGraw-Hill.
- Psacharopoulos, G., Tan, J. P., & Jimenez, E. (1986). *Financing education in developing countries: An exploration of policy options*. Washington, DC: World Bank, Education and Training Department.
- Psacharopoulos, G., & Woodhall, M. (1985). *Education for development: An analysis of investment choices*. New York: Oxford University Press.
- Rose, P. (2002). *Cost-sharing in Malawian primary schooling: from the Washington to the post-Washington consensus*. Unpublished doctoral thesis, University of Sussex.
- Sudharshan, C., & Xiao, Y. (2001). *Public health and education spending in Ghana in 1992–98: Issues of equity and efficiency* (Policy Research Working Paper, Vol. 2579). Washington, DC: World Bank.
- Thobani, M. (1983). *Charging user fees for social services: The case of education in Malawi*. Washington, DC: World Bank.
- UNESCO. (2002). *2002 education for all global monitoring report: Is the world on track?* Paris: Author.
- UNESCO. (2003, April–June). Education: Who pays? *Education Today*, p. 4–7.

- Wise, A. E. (1968). *Rich schools, poor schools: The promise of equal educational opportunity*. Chicago: University of Chicago Press.
- World Bank. (2004a). *Books, buildings, and learning outcomes: An impact evaluation of World Bank support To basic education in Ghana*. Washington, DC: World Bank Operations Evaluation Department.
- World Bank. (2004b). *Cost, financing and school effectiveness of education in Malawi*. Washington, DC: World Bank Development Research Group.
- World Bank. (2004c). *Summary education profile: Ghana*. Washington, DC: Author. Retrieved August 20, 2006, from <http://devdata.worldbank.org/edstats/SummaryEducationProfiles/CountryData/CtyDataSave.asp?SCtry=GHA,Ghana>
- World Bank. (2006). *Educational attainment database*. Washington, DC: Author. Retrieved August 20, 2006, from <http://www.worldbank.org/research/projects/edattain/edattain.htm>

# Globalization and Implementation of an Equity Norm in Higher Education: Admission Processes and Funding Framework Under Scrutiny

Gaële Goastellec  
*University of Lausanne*

Most of today's societies are confronted with an increasing necessity to legitimate the organization of their access to higher education. Commonly used as a yardstick to compare societies, the level of access to higher education is often presented as an indicator of the level of development and the capacity to produce knowledge, as well as a workforce adapted to the economic and social development. But increasingly, the issue is shifting from the outputs of general access to higher education to the specific institutions from which students gain admission. This raises the question of the fairness of higher education systems, their ability not to duplicate society but to produce social mobility, at least in the students' influx to and within the higher education sector.

By means of a collective research<sup>1</sup> dealing with national policies of access and equity in eight contrasted countries (Ethiopia, France, Ireland, Israel, South Africa, the United Kingdom, the United States, and Vietnam), fieldwork in South African and American institutions (interviews and participating observation), as well as a review of scientific literature, this study analyzes an international trend and its local variety of forms: the affirmation of an equity principle in the organization of access to higher education. This process is first perceptible in the evolution of

---

Correspondence should be addressed to Gaële Goastellec, University of Lausanne, Observatoire Science, Policies and Society, Head of Research Unit, Ch. de Bassenges 4, CH-1024 Ecublens VD, Switzerland. E-mail: gaele.goastellec@unil.ch

<sup>1</sup>This research took place within the Fulbright new century program 2005–2006. It was carried out by the Access and Equity working group, which includes Patrick Clancy, Heather Eggins, Sara Guri-Rosenblit, Phuong-Nga N'Guyen, Teshome Yizengaw, and me. Although this article draws on individual national case studies, the analysis developed here lies under the author's sole responsibility.

higher education admission norms, which changes gradually to take into account social identities when it comes to pondering academic results as a way to admit students. But the diffusion of this equity principle, which is defined as equality of opportunities, is also increasingly translated into higher education funding frameworks.

To scrutinize the implementation of the equity dynamic in access, this article is divided into two main sections. The first analyzes the historical changes in admission norms from a principle of “inherited” merit to an Equality of Opportunities principle. The second part of the article is concerned with the implementation of the Equality of Opportunity norm. It addresses this issue by looking at how higher education actors “traduce” (Callon, 1986), or “transcode” (Lascoumes, 1996) this norm into practices through two instruments: admission processes and funding policies. The higher education funding mechanism is also a reform of the higher education management (Johnstone, Arora, & Experton, 1998), providing institutions with a means to publicly account their provisions of equity in access. More broadly speaking, this article identifies a consequence of the globalization of higher education systems: the affirmation of the equity principle as a key point in the legitimating process of higher education organization and management. Thus, it attempts to analyze changes in the management of equity in access in the broader perspective of economy of inequalities (Piketty, 1997; Sen, 1999).

## NORMS OF ACCESS BETWEEN SOCIETAL ORGANIZATION AND HIGHER EDUCATION ROLE

The issue of equity in access to higher education is emerging on the political agendas of an increasing number of higher education public authorities and institutions’ governing bodies. This process can be analyzed as the consequence of three dynamics that globally weight on higher education systems: the demographic pressure, the economic pressure (which can be summarized as the diffusion process within the higher education systems’ concept of efficiency), and the political pressure (which calls for the diversification of the student body, especially when it comes to the selection of an elite; Goastellec, 2006a).

As a result, fairness in access is becoming an international standard and is therefore a determinant of higher education policies and comparisons. The conception and application of this new benchmark differs with each country examined in this study by specificities including size, structure, and origin of each nation’s higher education system. The comparison of the historical evolution of access to higher education in contrasted countries reveals three main periods, that is, three main norms successively constraining the organization of access. These norms reflect the transformation of a conception of both a legitimated social order and the role of the higher education systems within societies.



## INHERITED MERIT, LEGITIMATED INEQUALITIES AND MARGINAL ROLE OF HIGHER EDUCATION

The first identified norm has been conceptualized under the name of “inherited merit” (Clancy & Goastellec, 2007). Three dimensions characterize this norm, a principle component dating from the inception of the higher education system.

First, the initiators of the higher education institutions have a specific social origin: They are traditionally part of the elite. For example, in Europe, during the 13th and 14th centuries, the suzerains of the diverse kingdoms (such as Aragon, Castille, Leone, Portugal) established universities. More widely, in the 14th and 15th centuries, universities were created by political authorities and supported by religious leaders (Charle & Verger, 1994). In the following centuries South American religious communities implemented the first institutions, whereas in Indonesia it was the initiative of the Dutch colonizers. In each country—depending on the historical period—higher education institutions were created by an elite to answer specific purposes. As a result, the geography of the higher education systems used to echo the elite implantation.

The second characteristic concerns the geography of these systems, which are built on a highly centralized model that limits access to a restricted urban population: In South Africa, the first colleges were created in the 19th century in the Cape Province, where most of the British migrants were concentrated. In Indonesia, it was in Jakarta (formerly named Batavia) that the first faculties were set up. In France, their creation took place in Paris, and in the United States, the first colleges were established on the East Coast.

Besides this geographic limitation, both the goals devoted to the higher education institutions, such as the characteristics of the few disciplines taught (aimed at serving specific professions such as law and medicine in the first Parisian faculties), and the admission process (limited to the few high school graduates and most of the time constrained by institutional entrance examinations) increase again, the higher education degree of selectivity.

These dimensions underscore that the first higher education institutions were designed for young urban men coming from a small elite, although access could be marginally conferred to a handful of students from low-status families. At this stage, higher education participated in the reproduction of a minority's domination. Its role was to reinforce the power of an elite (often identified by a shared ethnicity, religion, profession, social status, or colonizers' position) and to favor the familial transmission of a few prestigious professions. For example, in the United States, the first eight colleges, which were founded before the American Revolution, were aimed at educating both the clergy and civic leaders (Lucas, 1994).

Nevertheless, although higher education systems are highly reproductive, they are so discreet in the national environment (as only a tiny percentage of an age group can access higher education) that they are not necessarily perceived as

greatly hampering social mobility. Be that as it may, access to higher education also reflects a restricted understanding of otherness: Some groups, the numerically dominant (such as women mostly everywhere, and non-Whites in South Africa), are considered as statutorily different, and thus they were not entitled to the same rights as the ruling group. The “inherited merit” also corresponds to the understanding of some social inequalities as legitimated, justified or “naturally fair.”

### EQUALITY OF RIGHTS: FORMALLY EQUAL THROUGH SEPARATED TRACKS

The implementation of the norm of equality of right can be analyzed both as a consequence of changes in the higher education roles and as a tool to understand the legitimacy of social organization. Modifications in the workforce, induced by transformations of the national economy, call for an increased access to higher education, whereas relief in major social conflicts, which had been structuring the society (regarding gender, ethnicity, religion, and/or socioeconomic background), encourages an opening of access to higher education for previously excluded groups. Indeed, the organization of primary education around the principle of equality progressively spread to higher education. We thus observe the diffusion of the ideal of universal access to primary education to the further steps of the education systems. Increasingly, this principle becomes fundamental in the organization of higher education, along with the ideal of meritocracy. Higher education remains elitist by principle and advocates a selective dimension of access. Many still consider academic performance to be a result of “natural intelligence” denying the influence of socioeconomic determinants on scholastic achievement, and so the influence of socioeconomic determinants is therefore denied.

As a result, equality of rights (or formal equality) has been implemented through the geographic decentralization of higher education and the diversification of higher education. This process took place within different periods. For example, in France, the number of universities first doubled between the end of the Second World War and 1970, before this dynamic expanded to midsize towns between 1980 and 2000 (Filâtre & Grossetti, 2003). In South Africa, new universities were built in the 60s and 70s, following the track of separated development (Waast & Gaillard, 2001). In Indonesia, each province was provided with a state university between 1956 and 1963. In Ethiopia, new regional universities were created in the 90s. The geographic development of the higher education sector also favors the integration of the system through the building up of national access policies. These policies can take the form of selection systems (such as the SAT in the United States or the UMPTN in Indonesia) or of legal norms regulating access (such as the national principle of equal access to universities in France).

This process of decentralization goes along with a diversification of the degrees proposed within universities and the creation of new kinds of institutions, usually aimed at providing nontraditional students with some higher or tertiary education (e.g., community colleges in the United States; colleges in Israel; IUT and BTS in France; private institutions in Ethiopia, Indonesia, and Vietnam). However, a shared rule consists of dedicating nonselective institutions to the enlargement of access.

Thus, the equality of rights in access is a legally sanctioned practice eclipsing former rules excluding some groups of the national population and formally through the opening of both new universities and other types of institutions. As a result, the former prerogatives of some ruling groups regarding access are preserved, whereas the demand for access of the other groups is partly answered. The result can be summarized as follows: formally equal, but apart.

### EQUALITY OF OPPORTUNITIES OR THE BUILDING UP OF SOCIAL PEACE?

The next step consists in the implementation of a norm of equality of opportunities. This dynamic is characterized by the shifting from access policies, focused on enlarging access, to a widening in the access. Indeed, statistical analysis of the students' influx toward the different higher education sectors bespeaks of the weight of social belongings on students' careers. For example, in France in 2005, students coming from employers' families represent 12.5% of the higher education students, but only 8.5% in preparatory classes to the "Grandes Ecoles," which lead to the most selective and prestigious degrees, compared with 16.6% in technological degrees (STS). Simultaneously, students originating from liberal professions or senior management families represent 31.2% of the students, whereas 51.9% are registered in preparatory classes to "Grandes Ecoles," with 13% in short technological degrees (MEN, 2005). We notice the same trend in the United States, where 25.5% of those from the richest families were, in 1999, to attend a highly selective 4-year institutions, with only 10.1% attending a 2-year public institution. By comparison, the students coming from the families with the lowest incomes were 5.8% to attend high selective 4-year institutions and 39% were to study in 2-year public institutions (McPherson & Schapiro, 2002). The fact that access to higher education has reached a universal level (more than 50% of one age group; see Trow, 1973) and that no more formal barriers limit access does not prevent the impinging of social background on access.

Several arguments can be identified to advocate the implementation of an Equality of Opportunity norm. The most general one questions the impact of enlarging the access (such as in the United States, Ireland, the United Kingdom, and France) of the structure of opportunities. Studies on this topic result in different

conclusions: Regarding the French context, Euriat and Thélot (1995) demonstrated a slight diminution of inequalities, whereas Goux and Maurin (1997) underlined a status quo, and others (Blöss & Erlich, 2000; Duru-Bellat, 2005) observe an increase in inequalities. Although the results differ according to the method used and the subsector analyzed, the probability is high—as no strong dynamic can be identified—that the enlargement of access does not profoundly affect the social structure of access. At the same time, higher education (particularly the most selective degrees) is the cornerstone of upward mobility. Even more, it becomes increasingly necessary to upgrade the obtained diploma to improve the level of membership in the status quo. The increasing access to higher education thus decreases the utility of the degrees in the marketplace and boosts the race for more higher education degrees.

As a result, with societies under constant scrutiny regarding their ability to organize themselves democratically (the development of New Information and Communication Technologies calling for a more detailed diffusion of information), it becomes increasingly important to provide data on access and to legitimize its role in the production of further economic inequalities. This political trend diffuses to higher education systems that are still highly elitists. It is, therefore, independent from the level of quantitative development achieved by the higher education system. In South Africa, for example, between 1994 and 1999, the political agenda shifted from the goal of increasing access to widen the student body to widening access at constant flux.

Another argument was probably the first to be used in the perspective of an Equal Opportunity norm; it deals with the compensation of former segregation. Unsurprisingly, it first appeared in the United States before emerging in other societies (such as Australia, South Africa, and India) in demand of remedial actions.

However, all these arguments bespeak a goal of legitimating the higher education organization and its effects, on the social structure.

First limited to a small number of institutions and systems, this norm is progressively spreading to an increasing number of institutions and higher education systems. The next part of this article analyzes how this norm has been codified by international bodies as a global target and implemented, in particular, through two instruments that denote local translations: admission processes and funding mechanisms.

## FROM NORM TO PRACTICES

### *Equality of Opportunity Normalization by International Bodies*

The Equality of Opportunities trend that we have previously observed has recently been the purpose of a normalization process at the international level. Indeed, if a norm is defined as a written document, resulting from a consensus



aimed at achieving an optimal level of organization in a specific context and applied on a voluntary basis (Borraz, 2004), equality of opportunities in access has become an international norm since it has recently been formalized by international bodies. For instance, the necessity to provide access to students “from disadvantaged backgrounds” was underlined in 2000 by the World Bank along with the UNESCO in the report “Peril and Promise: Higher Education in Developing Countries.” But the most important document regarding the normalization process is probably the “World Declaration on Higher Education for the Twenty-First Century: Vision and Action,” adopted by the 1998 World Conference on Higher Education. This report dedicates its third article to the question of Equity of Access. Two points of this article are particularly interesting, as they articulate two features of the equality of opportunities principle, although they can also contradict each other:

(a) In keeping with Article 26.1 of the Universal Declaration of Human Rights, admission to higher education should be based on the merit, capacity, efforts, perseverance and devotion, showed by those seeking access to it, and can take place in a lifelong scheme, at any time, with due recognition of previously acquired skills. As a consequence, no discrimination can be accepted in granting access to higher education on grounds of race, gender, language or religion, or economic, cultural or social distinctions, or physical disabilities.

(d) Access to higher education for members of some special target groups, such as indigenous peoples, cultural and linguistic minorities, disadvantaged groups, peoples living under occupation and those who suffer from disabilities, must be actively facilitated. (World Conference on Higher Education, 1998)

Several pieces of information stem from these extracts: first, the democratic rationality of the Equality of Opportunities norm, reference being made to the Universal Declaration of Human Rights. We observe here an enlargement of the problem of education being a basic right, from primary and secondary education to tertiary education. This democratic ideal can also be found in the discourses of those responsible for the organization of access at the local level, who justify the implementation of these admission processes, as one of the City University of New York community college admissions officers underscored:

Here one of our goals in the way we address the recruitment process is to convince black males to register and to attend classes, because where they come from it is not valued. And indeed, if you look at the student body, black men are a minority compared with black women. (personal communication, October 2005).

As I show later, depending on the localization of the higher education institutions, the democratic ideal is translated differently through a specific process aimed at compensating local or national inequalities in access.

Second, although higher education is considered a human right, it remains subordinated to a meritocracy principle. Third, both the prohibition to discriminate and the requirement to *actively* favor the access of minority groups are mentioned. *The articulation of these two rules* contains the seed of a strong contradiction that has already emanated from the now classical American debate of the 90s (D'Souza, 1993; Jencks & Philips, 1998; Sniderman, Carmines, Howell & Morgan, 1999; Thernstrom & Thernstrom, 1999), which questions the fact that by actively favoring minority groups, students with the best academic results were discriminated against.

The definition of this norm of access by international bodies, as we have already mentioned, does not have a constraining effect. Nevertheless, as a norm, it has a scientific, technical, and democratic legitimacy. It also has to be specifically translated into local practices. Access norms are translated into admission processes, which allow the regulation of the students' influx toward higher education and the different higher education subsectors and institutions. These admission mechanisms also coincide with funding mechanisms that sustain admission processes. Behind the Equality of Opportunities norm exists a variety of practices that correspond to a specific translation of this norm.

### *Equality of Opportunities in Admission Processes: Different Tools for a Same Goal*

Admission processes to higher education represent the hidden side of access. They consist in sociotechnique tools aimed at organizing the students' influx within higher education. A historical perspective reveals that the origins for ways to produce equality of opportunities in access goes back to the first half of the 20th century when the SAT was developed as a national entrance examination in the United States as a way to select the most intelligent individuals regardless of their social background. The aim was to produce a "classless society" (Conant, 1940). But the selection of individuals the SAT provided revealed collective inequalities in access to higher education (e.g., regarding socioeconomic and ethno-racial belonging) and reproduced social inequalities. Progressively, along with desegregation, tools to compensate historical disadvantages were implemented under the name of Affirmative Action. During the 90s, this formula was abandoned and holistic admission processes were organized (Goastellec, 2004) to measure the academic merit of an individual regarding all the handicaps he had to face to reach this level. This summary of the admission processes in the United States underlines the fact that the same goal can be pursued through different admission tools and, even more, that admission processes are permanently readapted when unexpected effects are isolated. We therefore observe an increasingly complex reading of

inequalities and, as a result, the equation of the Equality of Opportunities becomes still more cryptic.

As such, the Equality of Opportunity norm is implemented at different levels (national/institutional) and by different actors (public authorities/specific institutions).

Such an example can be found in Indonesia, where prestigious universities, concentrated on one island, have organized a second path of admission to enlarge the geographic (and thus social and ethnic) diversity of their student body (Goastellec, 2003a). In South Africa, the end of apartheid also increased the tension surrounding the admission process, leading to the use of affirmative action and a holistic process of admission by the more elitist institutions as well as an attempt to completely reorganize the admission process at the national level (Goastellec, *in press*).

In Ireland, in response to the statutory requirement set out in the Universities Act (Irish Minister of Education, 1997) and other legislation, most of the third-level colleges have introduced their own direct admissions procedures to deal with nonstandard admissions outside the framework of the Central Applications Office, which nationally allocates places to almost all higher education institutions on the basis of academic achievements. This pool of reserved places represents a form of affirmative action designed to facilitate access for students with disabilities, mature students, and students from socioeconomically disadvantaged backgrounds who would not meet the standard academic requirements for admission. In many cases the chosen affirmative action for school leavers from socially disadvantaged families consists in admitting students with levels of achievement which fall short of the Leaving Certificate point's requirements for traditional students (Clancy, 2006). In the same vein, the national higher education entrance exam used to admit students in Ethiopian universities incorporates some affirmative action dimensions by requiring a lower level of achievement from women, disabled students, and students from disadvantaged regions (Yizemgaw, 2006). In Israel, some of the universities also use forms of affirmative action to favor minorities' access (Guri-Rosenblit, 2006). In France, where universities are opened to all high school graduates, the promotion of minorities' access should be sought for among the selective *Grandes Ecoles*: During the last decade, the Parisian Institut of Political Sciences advertised itself as the flagship institution regarding these procedures by adopting a specific admission process for students coming from geographically identified as disadvantaged high schools (ZEP; Priority Area of Education).

Whether at the whole sector level or at the institutional level, the implementation of Equality of Opportunities is questioned nearly everywhere.

The comparison of these local processes reveals that every institution and higher education system is characterized by specific processes that are more or less formalized and codified. They also take place at the sector, institution, faculty, or department level regarding the culture or history of the institutions and higher

education system as well as its configuration (defined as the relationships between public authorities, institutions, and academic professions; Musselin, 2001). This comparison also shows that the characterization of legitimated identities that are taken into account to benefit from these equal opportunities processes differs regarding each society.

As a result, the analysis of universities belonging to distinct cultural areas shows extremely diversified databases. The statistical databases are built up to classify the population. They echo institutional practices and an original understanding of the Nation-State (Schor & Spire, 2005). More precisely, these categorizations depend on three factors: the historical construction of the nation-state, the specificity of the population's composition, and the tradition of public recognition of some specific identities. These categories are idiosyncratic: Each nation defines what is relevant according to its own history. In Indonesia, the geographic origin—the Province—is one of the tools used to classify the population. This statistical construction goes back to a specific understanding of identities, which is part of the national model of integration. Based on a heterogeneous geographic territory, the Indonesian State uses the regional origin to read social inequalities, whereas racial, ethnic, religious, and social origins are taboo and are restricted to the private domain. In the United States and in South Africa, it is the melting pot, the mix of races, and the ethno-racial indicator that have been the main criteria to measure social inequalities for a long time. In France, the Republican model consists of having the human being as a universal category. The measurement of social inequalities is done through socioprofessional categories, that is, a euphemized category of social classes. Far from being fixed, these information systems evolved depending on two factors: changes in society's social composition and its problematization, mainly regarding the negotiation between the State and the civil society.

The changes that admission processes undergo echo these national specificities and even the changes they sometimes anticipate. However, behind these differences, these databases express the way higher education institutions are legitimized in their relation with citizens and users. They also illustrate the democratic dimension of the universities' justification mode.

The understanding that inequalities of access is an issue is therefore central to the spreading of the norm of equality of opportunities as well as to the recognition of a broader social diversity. The comparative perspective shows that a certain "consensus" is about to be reached regarding the necessity of diversifying the identities that are taken into account in the admission process. As a result, the shared principle consists in developing more correspondence between admission processes and the complexity of the national social diversity. Henceforth, the equality of opportunity norm is accompanied by a deeper deconstruction of social belongings and by a more complex analysis of the way these belongings influence, in specific contexts, access to higher education and the rewards it provides.



### *Funding Equality of Opportunities*

Another instrument favors the implementation of equality of opportunity in access: It has been a long time since funding has been perceived as a tool to compensate socioeconomic handicaps, and thus to widen access. Although the cost-sharing rationale (mainly the split of the costs between public authorities, students, and families) is crucial on the agendas of an increasing number of higher education actors, this dimension is not summoned in this purpose. Much has been written about the international trend to make students bear part of the cost of their studies (see, e.g., Johnstone, 1986, 2002, 2006) through tuition fees, whether they are up front or repayable. Up-front fees are traditionally balanced by grants and loans aimed at widening access. However, these tools have a limited impact on the steering of access in institutions. They do not make institutions accountable for whom they register and for whom they graduate. Along with the added costs institutions have to bear when they register “at-risk” students, this explains why an emerging trend now consists in having public authorities using funding incentives toward institutions to reach this goal of widening access. Mainly, they index part of the institutional funding on the characteristics of the students they register, as well as the characteristics of those they graduate.

This international trend is in the process of being implemented in Ireland through a new funding model (2006–2008) that integrates State premium for an identified target group of students (Clancy, 2006). As a result, the State allocates funding based on the enrollment of designated groups. This is part of several funding principles, such as the three following examples: first, increasing opportunities for students from all types of backgrounds so that they benefit from higher education institutions; second, providing stability in funding to encourage efficiency/performance benchmarked against national and international best practices; and third, rewarding institutional responsiveness to national and regional needs (Higher Education Authority, 2006). In Ethiopia, since the 2003 Higher Education Proclamation, a new framework of funding equity is in the throes of implementation. It introduces a funding formula that takes into account the type of program, or course enrollment, of female and disadvantaged students (Yizemgaw, 2006). In Israel, although a funding formula is not directly involved, universities that use affirmative action to promote minorities’ access get budgetary assistance from the council of higher education (Guri–Rosenblit, 2006). The South African reform of the Higher Education systems is also an attempt to implement some form of equity funding incentives: the block grant received by each institution integrates indicators aimed at improving both institutional and student efficiency. Indeed, the equity dimension is intrinsic to the efficiency one: The policy of widening has limited impact if it is not pursued by a funding policy providing institutions with the means to graduate these students. In this perspective, the South African funding uses two indicators. On the one hand, the number of entering students

and the number of graduating students are taken into account and every institution is asked to reach a fixed national norm (22%). On the other hand, the proportion of “disadvantaged” students is also taken into account to increase the funding received. This new funding formula should allow the government to control the development of the system and incite institutions to plan their policies within periods of 3 years (through rolling plans) and to follow a policy of equity within an efficiency regulation. Last but not least, the French delegate ministry of higher education is also working on a similar funding framework.

The increased linkage between equity policies in access and funding policies demonstrates the innovation dynamics that are at play in the organization of higher education cost sharing or, more precisely, the inclusion of institutions in the cost-sharing rationale. This becomes increasingly possible because of the contract-based principle linking institutions to their public authorities and to the movement of international accountability. The comparison of these national processes thus reveals a common trend to rethink the equity funding, and the imbrications of the equity dimension with the efficiency one; institutions should then be evaluated regarding their ability to promote educational mobility. The institutional funding should become more equitable by taking into account institutional costs linked more to the students who need academic assistance than to traditional students.

## CONCLUSION

Admission processes and funding frameworks are not the only instruments used to implement Equality of Opportunities in access to higher education, nor have they yet achieved their transformations. Several methods are currently being utilized, starting at the first level of the education systems, and we are witnessing a permanent reinvention of tools aimed at widening access or at making it more fair.

However, admission processes and funding frameworks illustrate even more specifically two trends. First, they are the result of the local reinvention of the international norm through a translation process (Callon, 1986), consisting in the “production of meaning through the networking of autonomous actors and the transaction between heterogeneous perspectives” or a transcoding process (Lascoumes, 1996): Local and national actors transform the information they receive by aggregating spread positions, adapting practices, and so on. Finally, the transcoding process corresponds to the ability to build up public problems to make their steering possible. The evolution of the access norms to higher education reveals how societies reinvent and rethink themselves through the production and implementation of public policies. The history of the access norms underlines both the globalization of the higher education system, globalization being understood as the “process through which the production of global frames of interpretation of the world tends to escape Nation-State” (Muller, 2003), and the national and/or local “transcoding” of the new global referential.

Second, the responsibility of fair access is shifting, as public authorities are making institutions more accountable for what previously students were bearing the liability. This illustrates a double level of accountability: Institutions are under the public authorities' (and increasingly students') scrutiny to be more transparent regarding the role they play in (re)producing a social structure. At the same time, public authorities—and more largely nation-states—are evaluated in the international arena, among other sectors, according to the democratic window that their higher education systems represent. One of the new global referential thus characterizes higher education as a democratic warrant.

## REFERENCES

- Blöss, T., & Erlich, V. (2000). Les nouveaux acteurs de la sélection universitaires. Les bacheliers technologiques en question. [The new actors of University Selection: technology "french bac" graduates]. *Revue Française de Sociologie*, 41, 747–775.
- Borraz, O. (2004). Les normes, instruments dépolitisés de l'action publique. In P. Lascombes & P. Le Galès (Eds.), *Gouverner par les instruments* [Norms: depoliticized instruments of public action, Governing through Instruments] (pp. 123–161). Paris: Presses de Sciences Po.
- Callon, M. (1986). Elements pour une sociologie de la traduction. *L'Année Sociologique*, [Some elements for a Sociology of Translation. Domestication of the scallops and Fisherman of St-Brieuc Bay] 36, 173–206.
- Charle, C., & Verger, J. (Eds.). (1994). *Histoire des universités*. [The history of Universities] Paris: PUF, Que Sais-Je.
- Clancy, P. (2006). *Access and equity: National report of Ireland* (Fulbright New Century Scholars Program internal paper). Washington, DC: Council for International Exchange of Scholars.
- Clancy, P., & Goastellec, G. (2007). Questioning access and equity in higher education: Policy and performance in a comparative perspective. *Higher Education Quarterly*, 61, 136–154.
- Conant, B. (1940). Education for a classless society: The Jeffersonian tradition. *The Atlantic Monthly*, 165, 593–602.
- D'Souza, D. (1993). *L'éducation contre les libertés. Politique de la race et du sexe sur les campus américains*. [Illiberal Education: The Politics of Race and Sex on Campus] Paris: Gallimard.
- Duru-Bellat, M. (2005, September). *Democratisation of education and reduction of inequalities of opportunities: An obvious link?* Paper presented at the European Conference on Educational Research, Dublin.
- Euriat, M., & Thélot, C. (1995). Le recrutement social de l'élite scolaire en France. [The social recruitment of the school elite in France]. Evolution of inequalities from 1950–1990. Evolution des inégalités de 1950 à 1990. *Revue Française de Sociologie*, 36, 403–438.
- Filâtre, D., & Grossetti, M. (2003). La carte scientifique française. In M. Grossetti & P. Losego (Eds.), *La territorialisation de l'enseignement supérieur et de la recherche. France, Espagne, Portugal* (pp. 21–43). [The French Scientific map, The territorialisation of higher education and research. France, Spain, Portugal] Paris: L'Harmattan.
- Goastellec, G. (2003). D'un multiculturalisme l'autre: Les politiques universitaires et la justice sociale: une comparaison Etats-Unis-Indonésie. In G. Felouzis (Ed.), *Les mutations actuelles de l'Université* (pp. 109–129). [From one multiculturalism to another: higher education policies and social justice] Paris: PUF.

- Goastellec, G. (2004). Entre politique des quotas et égalité: L'Université de Californie à Berkeley. *Cahiers Internationaux de Sociologie*, [Between quotas policy and equality: UC Berkely] 116, 141–164.
- Goastellec, G. (2006). Accès et admission à l'enseignement supérieur; contraintes globales, réponses locales? *Cahiers de la Recherche sur l'Education et les Savoirs*, [Access and Admission to higher education: global constraints, local answers?] 5, 15–35.
- Goastellec, G. (in press). Nouvelle Afrique du Sud, nouvel enseignement supérieur: Entre équité et performance. *Education et Société* [New South Africa, New Higher Education: Between Equity and performance].
- Goux, D., & Maurin, E. (1997). Démocratisation de l'école et persistance des inégalités. *Economie et Statistique*, [Democratization of schooling and maintenance of inequalities]
- Guri-Rosenblit, S. (2006). *Access and equity: National Report of Israel* (Fulbright New Century Scholars Program internal paper). Washington, DC: Council for International Exchange of Scholars.
- Higher Education Authority. (2006). *Recurrent Frant Allocation Model*. Dublin: Author.
- Irish Minister for Education. (1997). *Universities Act*. Dublin: Author.
- Jencks, C., & Philips, M. (Eds.). (1998). *The black and white test score gap*. Washington, DC: Brookings Institutions Press.
- Johnstone, B. (1986). *Sharing the costs of higher education: Student financial assistance in the United Kingdom, the Federal Republic of Germany, France, Sweden and the United States*. New York: The College Board.
- Johnstone, B. (2006). *Financing higher education: Cost-sharing in international perspective*. Boston College.
- Johnstone, B., Arora, A., & Experton, W. (1998). *The financing and management of Higher Education: a status report on worldwide reforms*. Washington, DC: World Bank.
- Johnstone, B. (2002). Challenges of financial Austerity: Imperatives and Limitations of Revenue Diversification in Higher Education, *The Welsh Journal of education*, II, 18–36.
- Lascombes, P. (1996). *Rendre gouvernable: De la 'traduction' au 'transcodage'. L'analyse des processus de changement dans les réseaux d'action publique, CURAPP, La gouvernabilité*. [Making governable: From traduction to transcoding. Analysis of changes process within public action networks] Paris: Presses Universitaires de France.
- Lucas, C. J. (1994). *American higher education, a history*. New York: St Martin's Griffin.
- McPherson, M. S., & Schapiro, M. O. (2002). Changing patterns of institutional aid: Impact on access and education policy. In D. Heller (Ed.), *Conditions of access, higher education for lower income students* (pp. 73–94) Westport, CT: ACE and Praeger.
- MEN. (2005). *Repères et références statistiques*. [Statistical markers and References] Paris: DEP.
- Muller, P. (2003, November 4). *L'analyse cognitive des politiques publiques. Vers une sociologie politique de l'action publique*. [Cognitive analysis of public policies: Toward a political Sociology of public action] Paris, Communication au séminaire MESPI.
- Musselin, C. (2001). *La longue marche des universités*. [The long march of French Universities] Paris: PUF.
- Nguyen, P. N. (2006). *Access and equity: National report of Vietnam* (Fulbright New Century Scholars Program internal paper). Washington, DC: Council for International Exchange of Scholars.
- Piketty, T. (1997). *L'économie des inégalités*. [The Economics of Inequality] Paris: La Découverte.
- Schor, P., & Spire, A. (2005). Les statistiques de la population comme construction de la nation. In R. Kastoryano (Ed.), *Les codes de la différence. Race, origine, religion. France, Allemagne, États-Unis* [Population statistics as nation building. Codes of difference. Race, Origin, Religion: France, Germany, and the United States] (pp. 91–121). Paris: Presses de Sciences Po.
- Sen, A. (1999). *Un nouveau modèle économique, Développement, justice et liberté*. [Development as freedom] Paris: O. Jacob.



- Sniderman, P., Carmines, E., Howell, W., & Morgan, W. (1999). Essai sur différentes interprétations du facteur racial dans le système politique américain aujourd'hui. Analyse critique de *divided by color*. *Revue Française de Science Politique*, [A test of alternative Interpretations of the Contemporary Politics of Race: A critical examination of *divided by color*] 49, 265–293.
- Thernstrom, S., & Thernstrom, A. (1999). *America in Black and White, one nation, indivisible*. New York: Simon and Schuster.
- Trow, M. (1973). *Problems in the transition from elite to mass higher education*. Berkeley, CA: Carnegie Commission on Higher Education.
- World Bank. (2000). *Higher education in developing countries: Peril and promise*, Washington D.C.
- World Conference on Higher Education. (1998, October 9). *World declaration on Higher education for the 21<sup>st</sup> century: Vision and action*. Paris: UNESCO.
- Waast, R., & Gaillard, J. (2001). *Science in South Africa* (Country report). Institut for Research and Development, University of Stellenbosch, Stellenbosch, South Africa.
- Yizengaw, T. (2006). *Access and equity: National report of Ethiopia* (Fulbright New Century Scholars Program internal paper). Washington, DC: Council for International Exchange of Scholars.

# Aspects of Fiscal Federalism in Higher Education Cost Sharing in Latvia

Rita Kasa

*Stockholm School of Economics in Riga*

This case study explores devolution of low-income student subsidies, via the national student loans program, from the central to local governments in Latvia by means of decentralizing political and financial responsibility to provide public assistance to low-income students in obtaining funds for higher education. It describes municipal engagement in providing primary loan guarantees and in assuming full risk for low-income student loans. This article argues that although there are venues for local governments to support low-income students' access to higher education, the central government should sponsor this policy politically as well as financially.

Increased demand for higher education, and decreasing capacity of governments to tax, pushes governments worldwide to seek new ways of funding higher education (Johnstone, 2001; Johnstone, Arora, & Experton, 1998). In recent years there has been a dramatic, albeit uneven shift of higher education costs from predominantly government or taxpayers, to parents and students (Johnstone, 2003), with a forth party assuming a share of tertiary education costs being philanthropists (Johnstone, 1986). However, as governments seek new ways of delivering subsidies to students, shifts of costs appear to be occurring not only between taxpayers in general and citizens individually, but also between various levels of public administration. Specifically, the struggle over what level of government and to what extent it is responsible for funding students in higher education seems to transcend from being a characteristic of constitutionally federal countries to constitutionally unitary countries as well.

In many countries, both trends (i.e., cost sharing in higher education and fiscal decentralization) present new policy experiences. Study of these experiences form

knowledge to inform policy developments internationally, as policymakers seek new ways to address the issues of higher education finance and accessibility. This article presents a case study of devolution of student financial assistance, that is, allocation of political and fiscal responsibility to support low-income students from the central to local governments as an attempt to improve higher education accessibility for low-income students in Latvia.

Latvia is a constitutionally unitary country with two levels of public governance: the central or national, and municipal or local. Devolution of responsibility over funding low-income students from the central to subnational governments has been implemented via the national governmentally guaranteed student loans system, in effect since 2001. This system requires that students, in order to borrow governmentally subsidized loans, provide individual primary student loan guarantees in the form of a wage earning cosignatory, real estate, or securities. Orphans and children who have lost parent guardianship are the only students not required to provide primary student loan guarantees. Other students who cannot secure primary student loan guarantees themselves can ask municipalities to become their primary loan guarantors. Municipalities that have issued such student loan guarantees are effectively funding higher education via indirect subsidies enclosed in full risk of student default.

Municipal involvement in guaranteeing student loans in Latvia is an expression of fiscal federalism (Oates, 1972; Ter-Minassian, 1997) where local governments hold the decision-making power over whether to engage in this form of student support. The concept of fiscal federalism assumes that the provision of public services is determined largely by the demands for these services of the residents of the respective jurisdiction (Oates, 1972, p. 17). Supposing that there are low-income people in every local jurisdiction, one can question whether optional subsidization of students in the most vulnerable position to access higher education (i.e., economically disadvantaged students, delegated to second-tier governments) is a policy solution that supports equity in higher education access nationwide. One can also ask why municipalities should incur sizable liabilities in the form of student loan guarantees for something that does not guarantee benefits accruing to the particular local jurisdictions, as higher education is neither exclusively public nor nationally good (Hyman, 2002).

Guided by the propositions of the theory of higher education cost sharing (Johnstone, 1986) and the theory of fiscal federalism (Oates, 1972), this article analyzes decentralization of subsidies to low-income students in Latvia via the national student loans scheme. This article describes rationales that have motivated municipal engagement in guaranteeing student loans as perceived by local policymakers at respective jurisdictions. The purpose of this article is to evaluate, based on the current municipal experiences, feasibility of higher education cost-sharing models between the two levels of government, as it is implemented in Latvia via municipal guarantees to student loans. The article concludes with

recommendations for intergovernmental cost and risk sharing in granting access to student loans and higher education for low-income students.

## DESCRIPTION OF THE STUDY

The analysis of laws, regulations, governmental statistics, data provided by municipalities guaranteeing student loans and commercial banks that issue actual loans within the governmental student loans scheme, interviews with municipal policy makers, and local student aid recipients have all contributed to developing this article. Although all data were instrumental in arriving at the conclusions in this study, the primary data source has been semistructured interviews with 14 municipal decision makers at six municipalities in Latvia.

Several factors have determined the number of municipalities covered and the number of participants interviewed for this study. One factor is that municipal involvement in the national student loans program is recent and the total number of student loans guaranteed by municipalities is small. By the summer of 2005, when the data were collected, about 13% of local governments in Latvia had issued primary guarantees to 110 student loans. Municipalities engaged in this study had provided loan guarantees to 23 students. These local governments were purposefully selected for the study based on their size and location (large regional center vs. smaller urban area in peripheral location) and their economic wealth (contributor or beneficiary with the municipal Equalization Fund that redistributes funding generated by municipalities in 2004). The sample of municipalities was limited to urban jurisdictions to enable consistency in data analysis, because rural municipalities would present a different case for analysis.

Participants in this study were chosen through the process of internal selection (Bogdan & Biklen, 1998, p. 61), which took place during the fieldwork; selection of participants was based on their expertise and accessibility. When recruiting municipal decision makers (elected officials and local bureaucrats), potential participants in the study were either approached through a municipal spokesperson or contacted directly. In cases where municipal decision makers were contacted by the researcher directly, an agreement about the time and place for a face-to-face or telephone semistructured interview was reached right away. In instances where initial attempts to access study participants was made via contacting municipal spokesperson, local decision makers were approached as suggested by a spokesperson of respective local government. Informal contacts with municipal spokespersons were an advantage in this study as it allowed acquiring background information about issues on municipal involvement in subsidizing higher education students, and in some instances allowed for a better identification of local experts who later were invited to participate in the study. Overall, access to high-ranking elected and permanent municipal officials was gained. Factors determining the



number of officials interviewed per municipality in this study were expertise of informants and their availability within the time frame allotted for the fieldwork.

Interview data were categorically aggregated and analyzed in the context of the information provided by other relevant data sources, such as legal documents and statistical data. Several theoretical propositions about sharing costs of higher education between the two levels of government have guided data analysis.

Conceptual framework applied in this article outlines that subsidies to higher education allocated via student loans need to be targeted toward low-income students to enable their access to higher education (Johnstone, 2006). It also denotes that higher education equity funding responsibilities can be devolved to some extent because the informational advantages of local provisions maximize the efficiency of targeting these subsidies (Ahmad, Hewitt, & Ruggiero, 1997). At the same time, because higher education produces benefits to more than one local jurisdiction, costs of funding should predominantly occur at the central level. Furthermore, higher education subsidies to low-income students should be equally accessible across the nation, requiring a centralized supply of such programs.

## MUNICIPAL INVOLVEMENT IN HIGHER EDUCATION COST-SHARING VIA STUDENT LOANS

Cost sharing in higher education in Latvia, where students and their families are required to contribute to covering higher education costs, was officially introduced in 1991 with the passing of the Education Law. This law stated that the government funds higher education for only as many students as is necessary for satisfying national manpower needs. In more specific terms, although the central government does not limit the total number of students that public institutions of higher education admit every year, on the bases of the national planning it determines how many students can be admitted to study in each academic program at public institutions free of charge. Admission to these governmentally funded places of study is merit based; governmentally allowed number of students with highest scores is enrolled in state-funded slots in each academic program. Students with lower grades, who fail to enroll in state-funded places, can enroll for tuition set by the institution of higher education. This legislative move has transformed the higher education system in Latvia from an all-state-funded higher education into a dual-track tuition system (Marcucci & Johnstone, 2006), where students who are fully funded by the government study alongside those who pay the entirety of their tuition at public institutions of higher education. Contrary to public institutions, all students in private colleges have to pay tuition.

Although the government introduced tuition, adjustments in the student financial assistance system to aid students paying tuition were slow to follow. Even though the number of tuition-paying students rapidly increased from 32% in 1995

to 57% in 1998, and then to 77% in 2004 (Ministry of Education and Science, 2005), it was not until 1997 that a program of generally available (Johnstone, 2006) loans for covering student living expenses was introduced. It was not until 1999 that a program of generally available loans for covering tuition was introduced (See Table 1). In the scope of this student loans scheme, the government was paying from its budget not just the interest rate subsidy, coverage of a grace period, and some loan forgiveness, but also the actual loan amount. In 2001, the student loans program was reformed by involving commercial banks in actual lending to students while the government continued to provide a subsidized loan interest rate, grace period, loan forgiveness, and assumed its role as a secondary guarantor

TABLE 1  
National Student Loans Program in Latvia

Year	Loans for Covering Student Living Expenses	Loans for Covering Tuition	Year
1997	<i>Principal provider of funds:</i> The government. <i>Eligibility:</i> All non-failing full-time students at accredited HEI. <i>Annual interest rate:</i> Five percent. <i>Annual interest applied:</i> During studies and during the repayment period.  <i>Grace period:</i> Six months after graduation; three months after dropping out of HEI. <i>Repayment schedule:</i> Fixed schedule. <i>Debt forgiveness:</i> Based on the public manpower needs and social policy goals. <i>Eligibility for delayed repayment and frozen interest rate:</i> Borrowers in military service, on maternity leave, unemployed, students who continue pursuing higher academic or professional degree. <i>Securities required:</i> None.	<i>Principal provider of funds:</i> The government. <i>Eligibility:</i> All non-failing tuition-sponsored students at accredited HEI. <i>Annual interest rate:</i> Five percent. <i>Annual interest applied:</i> No interest rate during studies. Interest rate comes into effect one year after graduation. <i>Grace period:</i> One year after graduation; three months after dropping out of HEI. <i>Repayment schedule:</i> Fixed schedule. <i>Debt forgiveness:</i> Based on the public manpower needs and social policy goals. <i>Eligibility for delayed repayment and frozen interest rate:</i> Borrowers in military service, on maternity leave, unemployed, students who continue pursuing higher academic or professional degree. <i>Securities required:</i> None.	1999
2001	<i>Principal provider of funds:</i> Commercial banks <i>Eligibility:</i> All non-failing full-time students at accredited HEI.	All non-failing tuition-sponsored students at accredited HEI.	2001

(Continued on next page)

TABLE 1  
National Student Loans Program in Latvia (*Continued*)

<i>Loans for Covering Student Living Expenses</i>	<i>Loans for Covering Tuition</i>
<i>Annual interest rate paid by the student: Five percent</i>	
<i>Annual interest applied:</i>	
<i>During studies and during the repayment period.</i>	<i>No interest rate during studies. Interest rate comes into effect one year after graduation.</i>
<i>Exception: The government withdraws the interest rate subsidy from students who dropped out. These students are charged the interest rate set by the commercial bank that provided the loan.</i>	
<i>Grace period: One year after graduation. Three months after dropping out of HEI.</i>	
<i>Repayment schedule: Fixed schedule.</i>	
<i>Debt forgiveness: Based on the public manpower needs and social policy goals.</i>	
<i>Eligibility for delayed repayment and frozen interest rate: Borrowers in military service, on maternity leave, unemployed, students who continue pursuing higher academic or professional degree.</i>	
<i>Primary loan securities required: Students have to provide either a co-signatory, or real estate, or securities. Municipality can act as a primary guarantor for the student loan as well.</i>	
<i>Exception: The central government acts as a primary loan guarantor for orphans and students with no parent guardianship under age of 24.</i>	
<i>Secondary loan securities: The government guarantees 90 percent of the loan amount issued to students.</i>	
<i>Exception: The central government guarantees 100 percent of loans issued to orphans and children with no parent guardianship.</i>	
<i>Components of the central government's subsidy: (1) interest subsidy;</i>	
<i>(2) grace period and delayed loan repayment; (3) debt forgiveness;</i>	
<i>(4) secondary guarantee; (5) administrative costs.</i>	

*Note.* Table originally composed based on following governmental regulations on student loans: Cabinet of Ministers' Regulation Number 251 passed on July 15, 1997; Nr. 86 passed on March 12, 1999; and Nr. 220 passed on May 29, 2001.

for 90% of the loan amount to students. However, the availability of loans became restricted as borrowers were required to provide additional loan guarantees to be able to receive the loan.

In the new student loans scheme the central government has become a secondary loan guarantor because of the new requirement for student borrowers to provide individual primary student loan guarantees in the form of a wage-earning cosignatory, real estate, or securities. Only those students who are approved for the aforementioned primary guarantees are able to qualify for the governmentally subsidized higher education loans. The only group of students who do not need to

provide such primary guarantees to receive a loan are orphans and students with no parent guardianship under the age of 24, with the national government acting as primary guarantor or the risk bearer for loans to these students.

Nowhere in the current national student aid system are there provisions for subsidizing low-income students as a separate group needing financial assistance. Direct governmental student subsidies in the form of free tuition and monthly stipends are entirely merit based. The national government also does not assume any responsibility for supporting access to loans for low-income students who are unable to provide individual student loan guarantees. Instead, the central government has transferred responsibility of loan guarantees from students with economically disadvantaged backgrounds to the local municipalities.

Municipalities can provide primary student loan guarantees based on the municipal council's decision to launch such a local policy. By cosigning student loans, municipalities incur liabilities and assume full risk of student default, because loans for education have a higher inherent risk (Ziderman & Albrecht, 1995) than any other loan. Municipalities as primary loan guarantors for student loans become main bearers of risk. The central government, although it does subsidize loans, acts as a secondary loans' guarantor and acts only after it has been established through court procedure that the primary guarantor is unable to repay loans on which students who received these primary guarantees have defaulted. In the case of municipal primary loan guarantees, any and all student loan defaults are required to be covered by the local government's treasury.

Currently, there is no provision that would stipulate assistance of the central government to municipalities should they encounter defaults. This places a great pressure on the local budget. Thus, there is no real intergovernmental risk sharing in student lending between the two levels of the government. According to the current procedure, the local government must be declared by the court as unable to repay defaulted student loans for the national government to take on the debt and compensate the lender—commercial banks. Such a policy solution can hardly be classified as fiscally sound.

At the same time, the requirement for primary individual student loans guarantees is ill served to low-income students. This includes students whose parents are unemployed, underpaid, or paid only partially legally (by not declaring the full amount of payment so that employers may evade taxes), students whose parents are retired, students whose parents are disabled or have health problems that require considerable medical expenses, students older than 24 years of age whose parents are deceased or have their guardianship rights removed. Students who do not have parents with income sufficient enough to provide primary student loan guarantees for their children do not receive any support from the central government in accessing student loans. In this situation, municipal primary student loan guarantees is the only form of support that low-income students may receive to obtain student



loans and access higher education. Thus, municipal primary guarantees for student loans, if implemented, provide crucial assistance to low-income students in accessing higher education. Responsibility over enabling higher education access for economically disadvantaged people is being shifted here from the national to local governments.

There are no specific regulations set by the central government on eligibility criteria that local governments should apply guaranteeing student loans. The central government also does not stipulate what administrative procedures municipalities should follow. For student loan guarantees, municipalities do not need to seek the approval of the Council of the Municipal Borrowing and Guarantees, which has authority to either endorse or bar municipal loan guarantees. At the same time, the amount of guarantees for student loans counts toward the maximum amount of liabilities that municipalities are allowed to incur, which is 20% of the annual municipal budget. If the municipality exceeds this amount, it is considered unable to manage its liabilities any longer, and it then becomes subject to the fiscal stabilization process, which affects municipal ability to carry out new local projects. In sum, municipal guarantees to student loans are not directly monitored by national fiscal authorities. It is a municipality's responsibility to ensure that any liabilities incurred because of student loan guarantees do not negatively affect municipal fiscal standing.

Like the decision to implement local guarantees for student loans, formulation, financing, and administration of this policy is a municipality's responsibility. Further, this article describes what rationales have led local governments in this study to engage in providing primary student loan guarantees, what characteristics of these policies are across the municipalities, and what costs municipalities incur by implementing such policies as perceived by local policymakers.

## FORMULATION AND IMPLEMENTATION OF MUNICIPAL GUARANTEES TO STUDENT LOANS

Interviews with municipal officials in this study show that local decisions to provide municipal guarantees to student loans have often been taken in response to the requests of constituents and to equalize accessibility of higher education among residents of their jurisdiction. Constituents who are unable to provide required primary student loan guarantees have approached elected representatives, pleading for public assistance in accessing loans and higher education. As told by an elected official from the municipality with an unemployment rate of over 20%,

A mother who came to me as a municipal representative said that if I will not help her as a representative and if the municipality will not provide [a] student loan guarantee her child will not be able to continue her studies. Then she [the daughter] needs to

drop out because both the mother and the father are unemployed but the daughter studies at University of Daugavpils.

Although this quote illuminates the necessity for a need-based public assistance program, it also points out the political stakes for local decision makers who have the power to foster their electoral support by providing financial support to low-income students unable to complete or begin their university education because of financial reasons.

Yet political stakes for municipal decision makers are twofold. On one hand, there is the need to “at least minimize” (as one of the study participants puts it) the inequity in higher education access experienced by their constituents. On the other hand, municipal officials are cautious that extensive municipal engagement in guaranteeing student loans may fiscally damage the municipality. That would further result in other politically negative consequences.

To control both local political benefits and costs that may arise from municipally provided primary student loan guarantees, local governments engaged in this study have established a set of formal criteria that applicants for student loan guarantees must meet. Although each municipality is sovereign in formulating these criteria, because of some intermunicipal policy borrowing and, as it emerged from the data in this study, because of shared understanding about what policy tools could control the demand for student loan municipal guarantees, eligibility criteria for local primary student loan guarantees are similar in all local jurisdictions in the study.

In all municipalities studied for this research, applicants need to reside in a respective jurisdiction to qualify for local support. The next major factor for considering an applicant's eligibility is social welfare status of the student's family. Social welfare status is assigned by the very municipality—based on either the nationally established lower per-capita household income threshold or municipally established per-capita household income threshold, which cannot be lower than the nationally established one. It is the responsibility of municipal social services to verify whether a household qualifies for social welfare status. According to municipal requirements, applicants also need to have good academic standing. Further, an applicant must submit a personal statement or, at some municipalities, an official certificate verifying that no other individual primary student loan guarantees are available. There is also a request for various other documents that show applicants' academic persistence and postgraduate career plans.

At the same time, interview data with municipal officials in this study show that stated eligibility requirements for student loan municipal guarantees at most local governments are more elaborate than the ones that are actually implemented. One criterion that so far has not had any influence on selecting recipients of student loan guarantees is the applicant's chosen field. This criterion has been intended to target municipal support for students who study in fields of local manpower

priorities. Because there was no information about how many people would ask for municipal help in accessing loans and whether the municipality would be able to support them all, the criterion was also intended to decrease the pool of eligible applicants. As an official at the biggest municipality providing student loan guarantees explains,

There was a fear that our [municipal] resources will be too limited [to provide student loan guarantees to all applicants]. In that case we really could consider whether the city needs astronauts or we prioritize future teachers. At the moment we can provide [student loan guarantees] for astronauts as well.

To date, all municipalities included in this study have awarded local support for all applicants that had met criterion of need regardless of whether they qualified for other eligibility criteria set out by the local policy. Yet, because of the possible pressure on a municipal budget, local governments maintain a range of eligibility requirements, in addition to the applicants' income status, which serve as filters for reducing the number of applicants who would qualify for support. These various criteria also indicate that even though municipal officials are aware of need-based inequity in higher education access, municipal fiscal health is a more important issue for local governments than funding equity in higher education access. In other terms, should municipal officials consider that municipality no longer can afford providing assistance to all applicants for student loan primary guarantees, student financial need will diminish as a criterion for targeting local public assistance to higher education students. This leads to the argument that devolution of financial responsibility to fund equity in higher education from the central to local governments is not an effective way to promote education opportunity among economically disadvantaged groups.

## PERSPECTIVES ON REPAYMENT AND MUNICIPAL LIABILITIES BECAUSE OF STUDENT LOANS

Municipal involvement in providing primary guarantees to student loans within the national student loans scheme in Latvia is recent. From the municipalities covered in this study, the earliest municipal procedure on issuing student loan guarantees was passed in 2002, and the most recent one was just implemented at the time of data collection for this study in 2005. Therefore there was no hard evidence, as of yet, about the actual impact of liabilities because of student loan guaranteed on local budgets. Nevertheless, experiences from other countries could shed some light on the rate of student loan repayment which never reach the 100% mark (Ziderman & Albrecht, 1995).

Responsibility to repay defaulted loans lies upon its primary risk bearer, which in the case of locally guaranteed student loans in Latvia are municipalities. Although local officials do realize the probability of student default, there is rather limited understanding about the extent of costs that municipalities may incur because of this. Of all the local governments included in this study only one had some estimation about what percentage of student borrowers with municipal guarantees may default. This local government had projected about 20% default rate among those who will have received its primary student loan guarantees.

At the same time, officials interviewed at this local government perceived that there is no effective mechanism as to how to recover funds from delinquent borrowers. The strategy that the municipality had envisioned was to repay the full amount of the loan it had guaranteed to the lending bank as soon as there are indications that student borrowers do not comply with repayments, and then deal with defaulting borrowers later on. In this way, according to the local official, the municipality will save on the penalty payments to the bank, which are 0.1% of the amount borrowed per day of delay, as well as on the general annual interest rate assigned to the loan, which is 5%. The payments could be more if students dropped out of the university and the unsubsidized loan interest rate, which students have to pay in the case of dropping out, is higher.

Although all municipal officials talked about litigation against defaulting recipients, they were skeptical of success in achieving positive outcomes for the local government. Their skepticism was based on the transience of these students across the European Union and to other countries, as well as possible low lifetime income and unemployment of these borrowers.

Although local decision makers at six municipalities were aware of the potential negative effects that guaranteeing student loans may have on local budgets, their current assessment of these liabilities' impact was not a conservative one. When evaluating the fiscal burden incurred from student loan guarantees, local representatives compared the burden to other municipal liabilities assumed by providing for local infrastructure development investment projects. In light of such comparison, student loan guarantees seemed to be small with no significant effect on exhausting limits of municipal annual liabilities in the amount of 20% of annual local budget.

Missing in this analysis, however, was a perspective that liabilities for student loans are essentially expenses until the loan has been repaid. Furthermore, there is no clear local public benefit that a municipality would derive from assuming liabilities for individual students because returns from investment in higher education are exclusively neither local nor public. Based on uncertain local public benefits and an unknown fiscal burden to the local budget, it can be argued that municipal involvement in higher education cost sharing as full risk bearers for student loans is not fiscally efficient.



## RECOMMENDATIONS FOR INTERGOVERNMENTAL COST AND RISK SHARING IN HIGHER EDUCATION

This case study on municipal involvement in sharing costs of higher education via loan guarantees in the national student loans scheme suggests that disproportionate allocation of responsibility to local governments to formulate, finance, and administer policy on low-income student subsidies is not a feasible solution for equitable and efficient financing of higher education access. To maximize equity in higher education access while minimizing fiscal inefficiencies, there needs to be intergovernmental sharing of costs and risks in low-income student funding via student loans system as well as division of responsibilities in formulating and administering the policy.

The issue of which level of government is primarily responsible for assuming the primary risk of student default, and as a result a share of higher education costs for low-income students, is central to the policy. At the moment, all liabilities for guaranteeing student loans are assigned to local governments. The central government becomes involved as a secondary loan guarantor only if a municipality, as a primary cosignatory, is unable to repay loans in default that could cause bankruptcy. At the same time, because local governments are held responsible for student loan guarantees, municipal bankruptcy would diminish not only municipal ability to compensate loans in default but also its ability to continue providing this service to local constituents. An arrangement where municipalities bear disproportionately large responsibility, as compared to the central government, for low-income student loans is not a viable policy either for municipal fiscal health or for equitable distribution of need-based student subsidies.

To ensure that loans for higher education are available to all low-income students across the nation and that local governments who take the risk to cosign student loans do not suffer fiscal damage, the central government should incur primary costs of low-income student aid and provide a safety net for local governments to protect them from losses caused by defaults. One way the central government can do that is by ensuring municipally guaranteed student loans. In that way costs of student loan defaults would be covered by insurance "bought" by the central government and budgets of local governments would not suffer losses in real terms. Contribution on the part of municipalities under this system would be assuming liabilities of various lengths for national student loans. At the same time, to ensure that low-income residents of all local constituencies have access to these publicly guaranteed student loans it needs to be provided that all local governments participate in the system of student lending.

Provided that the central government finances the risk of municipally guaranteed student loans, formulation of policy on need-based student subsidies via national loans program should be assigned to the central level of government to

secure that public support to low-income students in accessing higher education is available nationwide. Leaving the allocation of assistance to the local governments does not guarantee implementation of such a policy across all municipalities. Each subnational government is an autonomous unit with varying local priorities and may not consider implementing primary student loan guarantees as necessary. Therefore the central government should request implementation of low-income student support on the local level via national student loans program. This request, however, should be accompanied with an educative campaign on the issue directed toward both municipal decision makers and the general public.

In terms of formulating eligibility criteria for governmentally insured municipal assistance, it should be provided nationally that all college-bound low-income students qualify for this aid regardless of their field of study and career aspirations. Municipalities, however, could maintain the discretion whether to apply the nationally or municipally set definition of low-income status. The case of Latvia shows that local governments predominantly apply the national definition of low-income or welfare status based on per-capita household income. At the same time, lower income threshold to qualify for welfare status set by municipalities is only modestly higher than the national one. Municipal social policies such as housing subsidies are based on the locally applied definition of low-income status. Therefore, allowing local governments to follow their definition of "low-income" would save on administrative costs of determining and verifying whether an applicant qualifies for this assistance on the basis of need.

This study indicates that municipalities are in a better situation to assess student need and income background than is the central government. Although there are issues of tax evasion in Latvia that inevitably will negatively affect accuracy of need assessment, local governments still have better information about their constituents and can provide a more accurate student need assessment than some centralized state agency. There is already a locally established network of social agencies to determine eligibility of students for need-based financial assistance. All municipalities in this study implement a need-based component in assessing students' eligibility for local support and utilize this network. To ensure a greater degree of transparency in providing need-based student aid, the municipal council should remain involved in approving student eligibility for local guarantees for student loans. Municipal councils in Latvia are locally elected and decisions of these local representative bodies are publicly available.

Although means testing of students eligible for national student loans subsidies can be assigned to municipalities, a responsibility of dealing with delinquent student borrowers who have received municipal support should be assigned to the central level of government. It would unify the effort funded by taxpayers to recover student loans as opposed to fragmented actions by each municipality (also funded by taxpayers) trying to ensure loan repayment from students who received locally guaranteed student loans. Having one agency responsible for recovering all

loans with a portion of national and local subsidy would save administrative costs and remove this administrative burden from local governments. It is also likely that more professional expertise in recovering loans would exist at the central level than in smaller municipalities.

Division of responsibilities between the central and local governments over formulating, financing, and administering loan subsidies to low-income students is possible in the scope of the national student loans system. Such division is a more feasible policy solution than a complete devolution of responsibilities over delivering financial assistance to low-income students nationwide.

## CONCLUSION

This case study on aspects of fiscal federalism in higher education cost sharing via student loans scheme in Latvia illustrates that it is possible to share higher education costs between several levels of government in constitutionally unitary countries. In Latvia, this is accomplished by assigning to local governments the primary risk for loans to low-income students, whereas the central government maintains its role as a secondary guarantor for student loans. Implementation of such a policy for municipalities, however, is optional. If they choose to implement it, they also set additional low-income student financial assistance eligibility criteria.

Although such a policy solution is possible, disproportionate assignment of responsibility to formulate, find, and administer need-based student aid to municipalities in the form of student loan primary guarantees from the central government can impair a municipality's fiscal viability, or it can end equity in access to low-income student subsidies. High risk of student primary loan guarantees is likely to deter local governments from passing or expanding such low-income student support policies. Furthermore, in instances when municipalities have passed such policies, there is little assurance that local budgets will not incur fiscal damage in real terms because of student loan defaults. To minimize negative fiscal effects and maximize availability of low-income student loan subsidies, the central government should be engaged in supporting this policy politically as well as financially.

## REFERENCES

- Ahmad, E., Hewitt, D., & Ruggiero, E. (1997). Assigning expenditure responsibilities. In T. Ter-Minassian (Ed.), *Fiscal federalism in theory and practice* (pp. 25–48). Washington, DC: International Monetary Fund.
- Bogdan, R., & Biklen, S. (1998). *Qualitative research for education: An introduction to theory and methods*. London: Allyn & Bacon.
- Hyman, D. (2002). *Public finance: A contemporary application of theory to policy*. Mason, OH: Thompson, South-Western.

- Johnstone, D. B. (1986). *Sharing the costs of higher education: Student financial assistance in the United Kingdom, the Federal Republic of Germany, France, Sweden, and the United States*. New York: College Entrance Examination Board.
- Johnstone, D. B. (2001). Those "out of control" costs. In P. Altbach, D. B. Johnstone, & P. Gumport (Eds.), *In defense of American public university* (pp. 144–178). Baltimore: John Hopkins University Press.
- Johnstone, D. B. (2003). Cost sharing in higher education: Tuition, financial assistance and accessibility in comparative perspective. *Czech Sociological Review*, 39, 351–374.
- Johnstone, D. B. (2006). Higher educational accessibility and financial viability: The role of student loans. In *GUNI Series on the social commitment of universities volume 1. Higher education in the world 2006: The financing of universities* (pp. 84–101). New York: Palgrave Macmillan.
- Johnstone, D. B., Arora, A., & Experton, W. (1998). *The financing and management of higher education: A status report on worldwide reforms*. Washington, DC: World Bank.
- Marcucci, P., & Johnstone, B. (2006). Tuition policies in comparative perspective: Theoretical and political rationales. In B. Johnstone (Ed.), *Financing higher education: Cost-sharing in international perspective* (pp. 55–74). Boston: Boston College Center for International Higher Education.
- Ministry of Education and Science. (2005). *Pārskats par Latvijas augstskolu darbību 2004. gadā: Skaitļi, fakti, tendences* [Annual report on Latvia's higher education in 2004: Numbers, facts, tendencies]. Riga, Latvia: Author.
- Oates, W. (1972). *Fiscal federalism*. New York: Harcourt Brace Jovanovich.
- Ter-Minassian, T. (1997). Intergovernmental fiscal relations in a macroeconomic perspective: An overview. In T. Ter-Minassian (Ed.), *Fiscal federalism in theory and practice* (pp. 3–24). Washington, DC: International Monetary Fund.
- Ziderman, A., & Albrecht, D. (1995). *Financing universities in developing countries*. Washington, DC: Falmer Press.



# Higher Educational Cost-Sharing, Dual-Track Tuition Fees, and Higher Educational Access: The East African Experience

Pamela Marcucci, D. Bruce Johnstone, and Mary Ngolovoi  
*International Comparative Higher Education Finance and Accessibility*  
*Project State University of New York at Buffalo*

Three universal demands characterize higher education globally: the demand for higher quality, for increased access, and for greater equity. In East Africa, where resources are highly constrained, no nation has been able to meet these demands on the basis of public expenditures alone. Instead countries have had to increase resources from nonpublic sources, including tuition fees. In countries with strong resistance to tuition fees and where the difficulty of taxation is combined with a daunting queue of competing public sector needs, a dual-track tuition policy is especially popular whereby the most capable applicants are financed from public resources and other qualified students are allowed admission on a fee-paying basis. This article studies dual-track policies in Tanzania, Kenya, and Uganda. We find that although rewarding ability, the dual-track policies did little to offer opportunities for the poor.

Public systems of higher education worldwide are caught between increasing public and private demand for their products, rising per-student costs, and flat or even declining governmental revenues. The public demand emerges from the increasing recognition of higher education as a major engine of national economic growth and provider of individual opportunity and prosperity. The private demand, or enrollment pressure—especially in Africa and other developing countries—begins in many countries with the sheer demographic increase in the traditional tertiary

---

This article was presented at the IREDU Conference on Economics of Education, Dijon France, June 2006

Correspondence should be addressed to Pamela Marcucci, SUNY at Buffalo, 468 Baldy Hall, Buffalo, NY 14260. E-mail: [pnm@buffalo.edu](mailto:pnm@buffalo.edu)

education age cohort, compounded by the increasing secondary school completion rates, which in turn increase the number of secondary school completers wanting to go on to higher education, further compounded by an expansion of what may be considered a college-going age cohort to include adults formerly bypassed by the system. The flat or declining governmental revenue—again, especially in most of Sub-Saharan Africa and other very low-income parts of the world—emerges from the poverty that not only leaves little wealth to be taxed but that also raises the opportunity costs of all public expenditures, which must compete with public sector needs such as elementary and secondary education, public health, public infrastructure, and other socially as well as politically compelling needs.

In response, most countries have turned to forms of private revenue supplementation for the support of their expanding higher educational needs—the most important of which is *cost sharing*, or the shift in higher educational costs from being borne mainly or even entirely by governments, or taxpayers, to being shared by governments, parents, and students (Johnstone, 1986, 2003, 2004a). The most important of these supplementary revenue streams, although not without problems and political resistance, are *tuition fees* paid for by parents (or larger extended families) and students themselves, mainly deferred or borrowed.

In Africa, donor-backed policies in the 1980s and 1990s de-emphasizing public expenditures on higher education relative to expenditure on elementary, middle, and secondary education contributed to a sometimes temporary, and sometimes not so temporary, reallocation of resources away from higher education and reinforced the call for more sharing of the costs of instruction by the students and families who benefit from it (World Bank, 1994; Ziderman & Albrecht, 1995). Donor pressure contributed to the capping in 1991 of Kenyan public university enrollments at 10,000 students per year with an annual growth rate of no more than 3% until 2017 (Kiamba, 2004), to the 17% reduction in government spending at Makerere University in 1991 (Ssebuwufu, 2003) and to decreases in government financial support to higher education in Tanzania in the late 1980s (Ishengoma, 2004).

A particular form of tuition fee policy that we have labeled *dual track* appears to achieve some real revenue supplementation but with problematic impacts on equity (Marcucci & Johnstone, 2007). Dual-track tuition policies are characterized by a highly restricted, “merit-based” entry to free or very low-cost higher education, with other applicants not so admitted permitted entry on a fee-paying basis. The origin of such a plan seems to lie in former Communist countries, in which free higher education was not only an expectation, frequently enshrined in a constitution or higher education framework law, but also where the country simply did not have sufficient tax revenue to accommodate all of the qualified applicants (Bain, 2001). The dual-track concept is especially popular in countries where strong resistance to tuition fees is to be expected, and where the difficulty

of taxation<sup>1</sup> combines with a particularly daunting queue of competing public sector needs, thus raising the opportunity costs of additional investment in higher education and raising the stakes on the quest for politically acceptable forms of revenue diversification.

This article is about a particular type of dual-track tuition policy in place in East Africa—Kenya; Uganda; and, until recently, Tanzania—and results from a study of East African tuition and access conducted by the University at Buffalo's International Comparative Higher Education Finance and Accessibility Project with the cooperation of the then vice chancellor of the University of Nairobi (now principal secretary in the Ministry of Science and Technology) and the Ford Foundation Office in Nairobi.

The research methodology included a review of policy documents and research studies, a research consultation in Nairobi, and surveys of students from the University of Dar es Salaam, Makerere University,<sup>2</sup> The Universities of Nairobi and Kenyatta, and St. Augustine University.

An obvious question raised by dual-track tuition policies is their impact on equity. Because entrance to the limited number of “free” places is by highly competitive examination, conventional academic wisdom would assume that the children of the well-educated and the privileged, with their access to the best secondary schools and all of the other advantages of their family cultural capital, would be disproportionately represented—even though these parents would almost certainly pay some tuition fees if necessary. The latter assertion is based in part on the rapid growth of tuition fee dependent private higher education as well as the popularity of the dual-track options in all of these countries.

The research undertaken by the Buffalo Project explored such questions as

- What is the difference in, for example, socioeconomic, geographical, and secondary school background between those who receive the free (or lower cost) places and those who must pay fees?
- How well do the cutoff examinations predict academic success?
- What is the profile of qualified students that decide not to opt for the fee-paying places?
- What is the money that is raised used for (e.g. salaries, increases in number of faculty, expanded capacity, filling in for further reductions in tax revenue)?

It must be noted that in the summer of 2005, Tanzania moved from a rather tentative dual-track tuition policy—wherein most students were entitled to free

---

<sup>1</sup>Not only are taxes difficult to collect, especially in developing and transitional countries, but the marginal net revenue—after the increasing costs of collection and the higher levels of escape and avoidance—begins to decline as the level of regressivity likely increases.

<sup>2</sup>The research at Makerere was undertaken in the context of dissertation research by Carrol (2004).

higher education, a moderate tuition fee was charged to the self-sponsored students, and modest fees were charged for food and lodging—to a policy in which all students pay a significant tuition fee, all tuition fees are deferred (as loans), and all food and lodging costs are deferred (also as loans).<sup>3</sup>

## DUAL-TRACK TUITION POLICIES IN EAST AFRICA

Following independence in the three countries, university students were entitled to free room and board, free tuition, and spending money (with the exception of Tanzania, where bursaries were only introduced in 1967). One of the rationales for such support was the expectation that most students would join their country's civil service after graduation to replace the departing colonial administrators. Political concerns may also have been behind such policies and leaders may have simply been looking for a place to park the potentially restive, politically charged, educated university-age cohort.

Therefore, government investment in education was highly skewed toward higher education. This situation began to change in the late 1980s when governments with the explicit encouragement of donors started to emphasize the importance of primary and secondary schooling for economic development and freeze or even decrease their relative investment in higher education. In Sub-Saharan Africa, public current spending on higher education as a percentage of total public current spending on education decreased from 19 to 16.7% between 1985 and 1995 and tertiary education expenditure per student decreased from 802% of the gross national product per capita to 422% (Task Force on Higher Education and Society, 2000). Whereas 17% of the World Bank's worldwide education-sector spending was on higher education between 1985 and 1989, 10 years later the proportion had declined to 7% (Bloom, Canning, & Chan, 2005). Moreover, these changes were taking place in a context of dramatically increased demand for higher education because of demographic growth and increased rates of secondary school participation.

Governments and university leaders in East Africa introduced dual-track tuition policies to expand higher educational capacity (and, they hoped, quality) despite these challenges without introducing politically unpopular tuition fees to all students and families.

In all three East African countries, the cutoff points for sponsored admissions are set based on government estimates of the number of students that they are able to support. Particularly in Kenya and Uganda, it is rapidly becoming more accurate to think of the university financing system as one in which most students have to pay tuition fees while only a few academically excellent students receive government sponsorship (Carrol, 2004).

---

<sup>3</sup>2006/07 Academic Year.



### *Dual-Track Tuition Policy in Uganda*

The dual-track policy was introduced in Uganda at Makerere University via the Private Entry Scheme (PES) in 1992 and later extended to all Uganda public universities (Court, 1999). Under the scheme, government-sponsored students do not have to pay tuition and they receive what in many African countries had been long considered a standard student entitlement: free room and board (it is probable that starting with the next academic year, even government sponsored students will have to cover their own room and board). The public universities run two different admission processes. The first, conducted by the Public Universities Joint Admissions Board (PUJAB), selects those students who will be awarded government scholarships (publicly sponsored students) based on the number of students that the Government of Uganda decides to sponsor. That number currently stands at about 4,000. Before the admissions process, all faculties within the universities provide information on the number of students that they can accommodate and decide on the distribution of government-sponsored and privately sponsored students.

All students who wish to apply for admission under government sponsorship are required to fill out the PUJAB application form in which they are asked to rank their top six choices of degree programs at public universities and four choices of diploma programs at other public tertiary institutions. The minimum qualification for entry into Makerere and other public universities is two principal passes at the Uganda Advanced Certificate of Education Examination. However, to earn a government scholarship, students need to be outstanding. Most students sit for either three or four subjects in their area of study (arts or sciences). Their scores on the various subjects are then weighted based on the requirement of individual programs within faculties, and the top-scoring students are admitted.

The cutoff point for admission into each program is determined by the lowest score of the last person accepted into that program. Very popular programs like medicine, dentistry, and architecture have high cutoff points, whereas the less popular such as law, mass communication, and social work and social administration have lower cutoffs. Affirmative action policies, which add additional 1.5 to 4 points to a student's scores, are in place for women, applicants with disability, talented athletes, and the biological children of Makerere employees (Carrol, 2004).

The second admissions process, for private admission, happens after the PUJAB admissions. Students who do not get a government scholarship are invited to put in applications under the PES. The private admission selection process is similar to the PUJAB process, and public universities do the admissions jointly. At Makerere, where programs are offered during the day and evenings, the higher performing students are put in the day programs, where they study together with the publicly sponsored students.

There are no legal limitations on the number of privately sponsored students that are allowed in the institutions and faculties differ in the proportion of private

students that they accept. However, government-sponsored students have first priority, and the Universities and Other Tertiary Institutions Act (The Act) of 2001 does give the board of an academic unit the power to regulate the admissions of students subject to the approval of the academic senate (Carrol, 2004).

The level of tuition fees for the private entry scheme students is set by the faculty subject to approval by the Academic Senate and the University Council. Fee levels vary and science faculties tend to charge more than humanities faculties. Tuition fees average about 1,800,000 Ugandan shillings (US\$994) per year.<sup>4</sup> Tuition increases are generally difficult to get passed by the University Council because of the government representatives who usually block such increases (Carrol, 2004).

Dual-track tuition policies have greatly expanded capacity as illustrated by the dramatic increase in enrollment at Makerere University between 1992 and 2002 from 5,000 to more than 30,000 students and by the growth in total public universities enrollments that reached 46,819 in 2004 (Ministry of Education and Sports, 2005). As of 2002, about 80% of the student body at Makerere was made up of privately sponsored students (Carrol, 2004). However, survey data (Carrol, 2004) suggest that the dual-track tuition policies do not increase the access of traditionally underrepresented groups given the absence of student financial assistance programs such as means tested grants and student loan programs and that, in fact, the private entry scheme may even reinforce existing inequities in participation at the university. There is little socioeconomic difference between the government and the privately sponsored students, with both coming from relatively affluent families. In the absence of a student loan program, enrollments are limited to those students whose families can afford to finance all the related costs of higher education.

Some students who do not qualify for government sponsorship apply to private universities or diploma granting institutions or go abroad rather than opting for the fee-paying places. Others who cannot afford the self-paying options try to raise money for admission at a later time or reapply the following year hoping to qualify for government sponsorship. Many opt to start working instead (Carrol, 2004).

Makerere University has generated large amounts of revenue from the private entry scheme: increasing from 4,080,059,201 Ush (US\$3,831,000) in 1995/96 to 29,438,099,000 Ush (US\$16,510,000) in 2002–03.<sup>5</sup> As shown in Table 1, by 2003, more than half of university funding was coming from the PES. It must be noted, however, that the revenues generated are not uniform over academic units. Some academic units that admit large numbers of private students have been able to raise significant amounts of revenue, whereas other units have not been as successful.

The bursar's office retains a portion of the tuition and fees that it collects and sends the remainder to the income generating units (faculties and institutes.) The

---

<sup>4</sup>US\$1 = Ush 1,810.3.

<sup>5</sup>1995–96 exchange rates US\$1 = 1,065 Ush and 2002–03 US\$1 = 1,783 Ush.

TABLE 1  
Privately Generated Funding at Makerere University, 1995–2003

<i>Year</i>	<i>Private Funding</i>	<i>Total Funding</i>	<i>% Private</i>
1995–96	4,080,059,201	24,408,492,201	17
1996–97	7,561,493,114	26,816,801,848	28
1997–98	8,799,261,213	28,299,261,213	31
1998–99	13,663,196,178	36,205,134,178	38
1999–00	15,080,261,764	38,070,261,764	40
2000–01	17,406,254,325	39,466,254,325	44
2001–02	19,030,439,000	45,680,439,000	42
2002–03	29,438,099,000	55,698,099,000	53

*Note.* Reported in Ugandan shillings. Source: Makerere University, Finance Department in Carrol (2004).

distribution amounts are set by the University Council and vary with the type of program (day, evening, etc.). The centrally held money is used for university-wide activities, such as supplementing staff salaries, supplying staff development, sponsoring research, and so on. Each faculty has some discretion over how it spends its own income subject to approval by the University Council, with over 90% going to recurrent expenditures and salaries (Carrol, 2004).

### *Dual-Track Tuition Policy in Kenya*

Very modest tuition fees were introduced in public universities in Kenya in 1991, but the generated resources were insufficient given the severely limited number of students. The Makerere model was introduced in 1998 via the self-sponsored, or Module II, programs.

The assumed average cost of each degree program is 120,000 Kenyan shillings (Ksh; US\$1,534)<sup>6</sup> per year of which the government covers 70,000 Ksh (US\$895) for the sponsored students (Module I) leaving the remaining 50,000 Ksh (US\$639) to the student to raise from the Kenyan Higher Education Loan board (HELB) or private sources. Governmentally sponsored students are entitled to a means-tested HELB loan that at best (and only for the poorest students) covers up to three fourths of educational and living costs for the year (maximum loan of Ksh 42,000 and maximum bursary of 8,000 Ksh; Otieno, 2004). HELB loans carry a 4% rate of interest and are repayable starting one year after completion of studies.

Students who attain the prescribed cutoff point are admitted into the regular state supported programs by the Joints Admissions Board (JAB), a non-statutory body made up of the vice chancellors, deputy vice chancellors, principals, and deans of

<sup>6</sup>2004 exchange rate US\$1 = 78.194 Ksh.

the six public universities and representatives from the Ministry of Education. In principle, Kenya Certificate of Secondary Education holders with C+ and better qualify for public university admission; however, this cutoff point depends on the total public university student capacity of about 10,000 students. Therefore, the JAB sets the entry cutoff for government-sponsored students from year to year. If a greater proportion of the students have high passes in a particular year, the cutoff will be higher and vice versa. For example, the cutoff for admission in 2005 admission was 64 points higher than in 2004 (Otieno, 2004). Although the basic cutoff score is required to qualify for government sponsorship, a student must also meet the subject cluster cutoff point to enroll in his or her chosen field.

Non-JAB students who are admitted on a self-paying basis gain entry to universities on the basis of different criteria that vary from university to university. At the very initial stages of the Module II programs, candidates had to be Form Four school leavers who met the minimum entry requirement of C+ but who did not meet the entry cutoff point for government sponsorship. In an attempt to increase the number of self-sponsored students, various institutions made admission conditions more flexible and accepted students from different academic backgrounds including holders of A-level certificates, Kenya Advanced Certificate of Education from the old 7-4-2-3 system, P1 primary school teaching certificate holders, diploma holders, and certificate holders from other governmentally recognized institutions (Otieno, 2004).

There are JAB students who turn down their places in the Module I programs and enroll in the self-paying program because they wish to finish their studies sooner (given the fact that all students enrolling in public universities are required to wait one year after they complete high school because of university capacity constraints) or because they were placed in academic programs that they have no desire to pursue.

In Kenya, tuition fees for privately sponsored students range from 96,000 Ksh (US\$1,227) for most programs to 450,000 Ksh (US\$5,754) for dental and medical programs.<sup>7</sup>

The Module II programs have been expanding since 1998 and have contributed to increased enrollments in public universities (Tables 2 and 3). Almost half of all students at the University of Nairobi are enrolled in these programs. In 2002–03 about 40% of all public university students were in the Module II programs (Table 3).

Survey data suggest that although students in both the Module I and Module II programs come from the better-off segments of society, a significantly greater proportion of the students in the Module II programs come from the richer segments and are concentrated in high- and middle-income families (89%) compared to students in the Module I programs (68%; Otieno, 2005).

---

<sup>7</sup>2004 exchange rate US\$1 = 78.194 Ksh.



TABLE 2  
Kenya: Increases in Enrollments  
in Public Universities

<i>Year</i>	<i>Enrollments</i>
1999–00	41,760
2000–01	42,346
2001–02	54,543
2002–03	59,593
2003–04	58,016

*Note.* Source: Kenyan Ministry of Education, Science and Technology.

These findings may be due to the fact that until recently the student loan program was available only to needy government-sponsored students in the Module I programs and students who attended private universities. Therefore, all privately sponsored students had to cover both their tuition fees and living costs without recourse to a subsidized student loan. However, in 2005, the Higher Education Loans Board negotiated with two commercial banks—National Bank of Kenya and the Cooperative Bank of Kenya—to lend money to students who had been admitted into universities locally, including the students who are enrolled in the Module II program. Once the students' places have been confirmed by the universities, the

TABLE 3  
Enrollment in Public Universities in Kenya by Track (2002–03)

<i>University</i>	<i>Enrollments</i>		
	<i>Regular</i>	<i>Module II</i>	<i>Total</i>
UoN	11,090	10,902	21,992
Moi	6,800	3,174	9,974
Kenyatta	7,200	8,856	16,056
Egerton	7,500	1,097	8,597
JKUAT	3,200	3,074	6,674
Maseno	4,300	1,231	5,531
WEUCST	700	—	700
Total	40,790	28,334	69,124

*Note.* UoN figures are from Kiamba (2004) for 2002–03 academic year; the rest of the figures apply to the 2003–04 academic year. Source: HELB, courtesy of Public Universities (Otieno 2004). UoN = University of Nairobi; JKUAT = Jomo Kenyatta University of Agriculture & Technology; WEUCST = Western University of Science and Technology.

banks pay from 50,000 Ksh to 5,00,000 Ksh directly to the universities depending on the tuition costs of each academic program. The commercial bank loans have higher interest rates (14–15%) than the government loan (4%). It is not yet clear how much the enrollments have increased as a result of commercial bank lending to students.

In Kenya, there is tension between the Module I and Module II students. The government-sponsored students often view the self-sponsored students as unqualified and allowed to study only because they can afford to pay. Students argue that facilities are not adequate to accommodate such a large number of fee-paying students (Otieno, 2004).

The importance of the Module II programs as revenue earners has been growing since their introduction. In 1997–98, the Module II programs at the University of Nairobi generated about 4% of its total income, by 2002–03, this had grown to 33% (Kiamba, 2004). In the 2002–03 academic year alone, the University of Nairobi earned US\$17,551,873 through its parallel programs, and by the end of that year, income from students and parents (including both Module I and II) contributed close to 40% of the total university income (Otieno, 2004). In turn, the government allocation dropped from 70% of the university's income in 1995–96 to 49% in 2002–03 (Kiamba, 2004). Table 4 shows various income-generated activities.

The income from the Kenyan parallel programs is used for institutional development and payment of academic and administrative staff. Generally, 35% of the raised funds are used to pay the lecturers, whereas 65% goes to the university. Funds are used for improved teaching materials and building projects. The University of Nairobi is reported to have spent well over 520 million shillings on renovation and completion of stalled building projects using money from the Module II programs.

TABLE 4  
Income Earned From the Various Income-Generating Activities Through  
UNES, 1997–2002

<i>Year</i>	<i>Module II Programs</i>	<i>Other Projects</i>	<i>Total</i>
1997–98	12,964,110	66,696,046	79,660,156
1998–99	233,153,499	82,001,499	315,154,998
1999–00	377,144,631	84,160,615	461,305,246
2000–01	602,836,675	78,166,941	681,003,616
2001–02	944,096,451	73,359,334	1,017,455,785
2002–03	1,209,512,592	106,877,915	1,316,390,507
Grand total			2,870,970,308

*Note.* Reported in Kenyan shillings. Source: University of Nairobi (2003) in Kiamba (2004). UNES = University of Nairobi Enterprises and Services Limited.

TABLE 5  
Tanzania: Undergraduate  
Enrollments in Public Universities

<i>Year</i>	<i>Enrollments</i>
1995–96	8,350
1996–97	9,370
1997–98	10,773
1998–99	12,069
1999–00	12,665
2000–01	13,987
2001–02	15,047

*Note.* Source: Ishengoma (2004).

### *Dual-Track Tuition Policy in Tanzania*

In Tanzania, while foreign and institutionally supported students started to be admitted on a fee-paying basis in the early 1980s, the explicit dual-track tuition policy was introduced in a context in which cost sharing was already underway in higher education. In 1992, students (and families) became responsible for paying for their own transportation, application, registration, entry exam, and student union fees as well as caution money, and in 1993 student allowances were eliminated. In 1996, the University of Dar es Salaam's Council approved an official proposal for admitting privately sponsored Tanzanian students, and in 2002 it officially recommended that the university fill remaining spots not filled with government-sponsored students (who did not have to pay tuition fees) with privately sponsored, tuition-fee-paying students. In the same year, it voted to give

TABLE 6  
Income Generated From Private Tuition Compared to Government Investment at the  
University of Dar es Salaam

<i>Year</i>	<i>Private Tuition</i>	<i>Total Income</i>	<i>Private Tuition as % of Total</i>
1995	41,898,950	4,585,030,348	0.9
1996	78,285,199	6,582,493,050	1.2
1997	327,407,317	7,959,722,061	4.1
1998	393,755,289	7,155,431,989	5.5
1999	273,691,653	9,545,694,105	2.8
2000	611,977,434	11,163,220,908	5.4
Total	1,727,015,843	46,991,592,460	3.6
	US\$3,795,138	US\$103,264,607	

*Note.* Source: Ishengoma (2004).

the sons, daughters, and spouses of university staff and members of the University Council the right to pay only half of the tuition fees (Ishengoma, 2004).

The dual-track tuition policy in Tanzania was essentially changed when the government introduced student loans (July 2005) for the 2005–06 academic year to cover tuition fees, other academic fees, and room and board for all higher education students whether government or privately sponsored in the public universities or self-paying in the private universities. This student loan policy dramatically changed the country's tuition policy, moving it from a dual-track policy to one in which all students must pay tuition, albeit largely deferred as a loan to be repaid once they have finished their studies.<sup>8</sup>

Under its dual-track policy, the University of Dar es Salaam established criteria and set minimum cutoff points for admission in the individual degree programs that were based on the number of students that the government set for admittance under its sponsorship. Unlike Kenya and Uganda, the government also determined the distribution of students among campuses and programs (Ishengoma, 2004).

Admission to the government-sponsored places was based on pass mark achievement on the Advanced Certificate of Secondary Education Examinations. The minimum entry cutoff points set by the University of Dar es Salaam ranged from 6.5 to 10.5 points depending on the degree programs, with female applicants having a slightly lower cutoff point to make up for past discrimination. A limited number of nontraditional students entered public universities through Mature Age Entry Examinations and through distance learning conducted by the Open University of Tanzania that operates in all 25 regions of Tanzania Mainland.

Admission to the self-sponsored places was also based on results of the Advanced Certificate of Secondary Education Examinations exam. Candidates had to receive principal-level passes in appropriate subjects with a total of at least 5 points from three subjects obtained at the same sitting. Like in the other two countries, the different programs had additional admission criteria. Tuition fees for the privately sponsored students ranged between 600,000 Tanzanian shillings (Tsh; US\$550) and 1,00,000 Tsh (US\$917).<sup>9</sup>

In Tanzania, although there has been a significant increase in undergraduate enrollments in the past 10 years (see Table 5), with the University of Dar es Salaam alone growing from 3,146 students in 1993–94 to 14,221 students in 2003–04 (Bloom et al., 2005), only a small part of this growth (see Table 6) has been in self-sponsored students<sup>10</sup> despite increased applications and increases in

<sup>8</sup> At present (2007) there appears to still be a distinction between government sponsored students who pay a lower (largely deferred) tuition fee and privately sponsored students who pay a higher (but also largely deferred) tuition fee.

<sup>9</sup> 2004 exchange rate: US\$1 = 1,089.33 Tsh.

<sup>10</sup> Although the number of privately sponsored students grew from 106 in 1992–93 to 289 in 2001–02, it remained at about 3% of total enrollments throughout the period (Ishengoma, 2004).



the number of secondary school leavers (Ishengoma, 2004, 2006). As part of the dismantling of the dual-track tuition policy, the University of Dar es Salaam encouraged privately sponsored students to switch to government-sponsored places via a public announcement on their Web site, as long as they have a good reason. Given the movement from a dual-track tuition policy to general tuition (albeit deferred) for all, it was not possible for the recent research to gather information on the socioeconomic background of the self-sponsored students. However, it was observed that in general in Tanzania students from better-off families disproportionately undertake higher education (Johnstone, 2004b). It will be important to track the impact of the new tuition policy and loan program on the socioeconomic composition of the student body.

Given that enrollment in dual-track tuition programs was very low, relatively little additional income was generated. Between 1995 and 2000, only 1,727,015,842 Tsh (US\$1,583,393) was raised in private tuition (Ishengoma, 2004). The generated income was largely used to top up the salaries of the faculty who taught courses in the dual-track program (Ishengoma, 2004). The new policy would seem to be a significant increase in cost sharing, but only if the new tuition fees are recovered.

### EAST AFRICAN DUAL-TRACK TUITION POLICIES AND THE EQUITY OF HIGHER EDUCATIONAL ACCESS

The theoretical relationship between a dual-track tuition policy and the equity of higher educational participation—defined as the degree to which higher educational participation is correlated with socioeconomic class (or ethnicity, gender, language, or region)—is complex and depends on three interrelated factors:

1. *The total additional revenue made available to the higher educational system by virtue of the dual-track tuition fee policy:* The degree to which the revenue from the dual-track tuition fee payers can be captured by the higher educational system as opposed to merely enabling governmental budget cuts (or going into other wasteful or corrupt expenditures). Clearly, for a dual-track policy to lessen rather than to aggravate participation disparities, there first has to be some net additional revenue. In the East African context, the additional revenue was retained by the institutions and contributed to improvements in facilities, additional staff and better staff morale through higher salaries.
2. *The additional higher educational capacity made possible by the additional revenue:* The greatest contributor to inequitable participation is arguably the lack of capacity and the resulting need for stringent selection. As long as the current system of selection yields a disproportionately middle and upper income student population—which is inevitable in any system of selection

based mainly on measures of academic preparedness or even academic ambition—virtually any increase in capacity is likely to lead to at least a slight increase in the equity of higher educational participation. (That is, the sons and daughters of the wealthy can be assumed to have previously been “taken care of,” if not by the limited governmentally sponsored places, then by the private sector or by being sent abroad for their higher education.) In the three countries of East Africa, and particularly in Uganda and Kenya, university capacity grew significantly as a result of the dual-track tuition policy; what changed little was the socioeconomic background of the students.

3. *The provision of additional student financial assistance—in the form of means-tested grants or student loans—to low-income students who are qualified but only for the dual-track entry that is made possible by the additional revenue:* Some of the additional revenue, in addition to making possible the necessary increased capacity referenced previously, must make it more possible for the government to provide additional assistance that increases the participation of students of students who would now—by virtue of the additional capacity—be included but who would likely be unable to participate because of the low incomes of their families. There is no evidence in any of the countries that the additional revenue generated by the dual-track tuition policies was used to address equity concerns.

### SOME CONCLUSIONS ON THE SUCCESS OF THE POLICIES IN EXPANDING CAPACITY AND QUALITY AND INCREASING PARTICIPATION AND EQUITY

Based on the findings of the research, we conclude the following:

1. The dual-track tuition policy has had a beneficial effect on the financial viability certainly of Makerere and Nairobi, and it is presumed to have had a somewhat positive impact on the University of Dar es Salaam, Kenyatta University, and other higher educational institutions where it has been introduced.
2. The willingness of parents (and extended families and others) to contribute toward the higher education of those who are attending on a fee-paying basis is a strong indicator that many of those now being admitted to the universities on governmental sponsorship, or the tuition free basis, would pay a modest tuition fee.
3. Although the additional revenue streams from the privately sponsored students increased institutional viability and expanded capacity, they expanded access only for middle-income students and not for genuinely poor students

- because of the absence (until recently in Kenya) of means-tested loans available to privately sponsored students.
4. The successful Kenyan Higher Education Loan Scheme should be continued with minimal subsidization and maximum efforts to recover loans in default or arrears. Uganda should implement a similarly effective loans scheme as an autonomous public corporation. Attention should be given in Tanzania to collection mechanisms and means testing.
  5. As soon as a “track record” of collection and loan recovery has been established by the public agency, efforts should be made to securitize the loan notes and thereby to release the government from some of the burden of initial capitalization entirely from its operating budget. As the loan notes become able to be privately capitalized (and thus no longer entirely a drain on the government’s operating budget), eligibility to borrow should be extended to students in the privately sponsored tracks who cannot attend without some form of financial assistance. Criteria for borrowing should continue to be a combination of academic promise (however measured) and financial need—that is, assuming an expected family contribution.
  6. Too little is known about the academic success of those students who enter on a privately sponsored basis—and especially about the differences between the governmentally and the privately sponsored students. Research with some empirical evidence is needed to counteract (or validate) assumptions or rumors about the academic worthiness of the fee-paying, or privately sponsored, students.

## REFERENCES

- Bain, O. (2001). The costs of higher education to students and parents in Russia: Tuition policy issues. *Peabody Journal of Education*, 76, 57–80.
- Bloom, D., Canning, D., & Chan, K. (2005). *Higher education and economic development in Africa*. Washington, DC: World Bank (AFTHD).
- Carrol, B. (2004). *Dual tuition policy in Uganda*. Unpublished research report prepared for the International Comparative Higher Education Finance and Accessibility Project.
- Court, D. (1999). *Financing higher education in Africa: Makerere: The quiet revolution*. Washington DC: World Bank.
- Ishengoma, J. (2004). Cost sharing in Tanzania: Fact of fiction? *Journal of Higher Education in Africa*, 2, 101–134.
- Ishengoma, J. (2006). *Students’ backgrounds, higher education costs and attitudes toward cost sharing in higher education in Tanzania: Case study of the University of Dar es Salaam and St. Augustine University of Tanzania*. Unpublished research report for the International Comparative Higher Education Finance and Accessibility Project.
- Johnstone, D. B. (1986). *Sharing the costs of higher education: Student financial assistance in the United Kingdom, the Federal Republic of Germany, France, Sweden, and the United States*. New York: College Entrance Examination Board.

- Johnstone, D. B. (2003). Cost-sharing in higher education: Tuition, financial assistance, and accessibility. *Czech Sociological Review*, 39, 351–374.
- Johnstone, D. B. (2004a). The economics and politics of cost sharing in higher education: Comparative perspectives. *Economics of Education Review*, 20, 403–410.
- Johnstone, D. B. (2004b). Higher education finance and accessibility: Tuition fees and student loans in Sub-Saharan Africa. *Journal of Higher Education in Africa*, 2, 11–36.
- Kiamba, C. (2004). The privately-sponsored students and other income-generating activities at the University of Nairobi. *Journal of Higher Education in Africa*, 2, 53–74.
- Marcucci, P. N., & Johnstone, D. B. (2007). Tuition policies in a comparative perspective: Theoretical and political rationales. *Journal of Higher Education Policy and Management*, 29, 25–40.
- Ministry of Education and Sports. (2005). *Education sector annual performance report 2005*. Dar es Salaam, Tanzania: Ugandan Ministry of Education and Sports.
- Otieno, W. (2004). *The private entry university scheme in Kenya*. Paper presented at a Consultative Workshop on Dual Track Tuition in East Africa. Nairobi, January 2005.
- Otieno, W. (2005). *Dual track tuition in Kenya's public universities: A study of the circumstances and conditions under which student attitudes towards cost sharing would change*. Unpublished research report for the International Comparative Higher Education Finance and Accessibility Project.
- Ssebuwufu, J. P. M. (2003). University financing and management reforms: The experience of Makerere University. In B. L. M. Mwamila, I. Omari, & E. Mbuya (Eds.), *Financing of higher education in eastern and southern Africa: Diversifying revenue and expanding accessibility* (pp. 85–95). Dar es Salaam, Tanzania: University of Dar es Salaam.
- Task Force on Higher Education and Society. (2000). *Higher education in developing countries, peril and promises*. Washington, DC: World Bank.
- World Bank. (1994). *Higher education: Lessons of experience*. Washington, DC: World Bank.
- Ziderman, A., & Albrecht, D. (1995). *Financing universities in developing countries*. Washington, DC: Falmer Press.



# Do Electronic Technologies Increase or Narrow Differences in Higher Education Quality Between Low- and High-Income Countries?

Norman Clark Capshaw

*Army Evaluation Center, Alexandria, VA and the University of Phoenix*

The disruptive technologies of the Internet and computers are changing our world in myriad ways. These technologies are also increasingly being employed in higher education but to what effect? Are the effects on higher education quality measurable, and if so, what is the effect on the traditional gap between high-income and low- to middle-income nations on this score? This theme is pursued in this article, which uses a variety of methods to probe the question. Because great controversy attends the notion of institutional quality, measures differ, and the effect of these technologies on that quality depends to a great extent on the definition being used. Low- to middle-income countries' usage of the Internet and computer technologies lags behind that of high-income countries, but projections indicate they are catching up.

Internet and computer technologies have become an important part of higher education, not only in the United States and other high-income countries, but increasingly so in so-called developing countries as well. Do these technologies make a difference, and if so, will they enable a closing of the gap in higher education quality that traditionally follows the gap in income between the rich and poor countries? This question is addressed by first examining the experience of the Internet and computer technologies in higher education in the United States, then turning the focus to the experience of higher education institutions in low- to middle-income countries.

Focusing on higher education in the United States, a recent Pew Research (Jones, 2002) report noted that 86% of college students have been online, 89% of

those students have a positive view of the Internet, and *almost 80%* feel that the Internet has had a positive impact on their higher education experience. A majority of the same students subscribe to academic-oriented Internet mailing lists and use e-mail to communicate the instructors and with each other concerning assignments. With such large numbers, one might be led to expect that it would be quite simple to demonstrate that computers and the Internet are contributing positively toward the quality of higher education.

### INSTITUTIONAL QUALITY AS MEASURING VIA "BEST OF" LISTINGS

Part of the challenge of determining the impact of Internet and computer technologies on higher education is arriving at a consensus definition for *quality*. Institutional quality is an elusive concept. Although many observers have an intuitive understanding what it entails, enumerating its essential elements can be a difficult and controversial task. One such measure that is highly controversial, yet always highly anticipated, is the *U.S. News and World Report* (<http://www.usnews.com>) annual listing of America's best universities. To almost no one's surprise, institutions such as Harvard, Columbia, Yale, Princeton, MIT, and Stanford appear at or near the top of this list every year. Likewise, there does appear to be some consensus even among academics that these institutions are among the best. But beyond the apparent consensus about the very top institutions comes great controversy about how other institutions fare under *U.S. News's* scrutiny.

*U.S. News* artfully deflects this controversy by focusing the attention on the criteria chosen for its prioritization. In its 2003 assessment of the top U.S. undergraduate institutions, *U.S. News* included the following nine criteria in its analysis: a peer assessment score, student graduation and retention rates, class sizes, student:faculty ratio, faculty resources, percentage of faculty who are full time, student performance on SAT/ACT tests and high school ranking, institution financial resources, and alumni giving. The publication responds to its critics by modifying these criteria from year to year.

In developing the list, *U.S. News* also compiles statistics on the computer and Internet availability to students at the universities chronicled in its annual list. One of the data listed is the number of library volumes available at each institution. In addition, the list includes the number of computers available to students at the schools. It might be expected that, as the number of library volumes and the number of computers available to students increases, the ranking of the school would also climb. One would therefore expect a *negative correlation*<sup>1</sup> between these variables and the institution's rank.

---

<sup>1</sup> Because a smaller number denotes a higher ranking school.

TABLE 1  
Correlation of Institution Ranking With  
Library Volumes

<i>Variable</i>	<i>Rank</i>
Library volumes	-0.403***
Volumes/student	-0.524***

\*\*\* $p = .01$ .

This is indeed the case with the number of library volumes and volumes per student. The strong negative correlations indicate that there is some relationship between the number of library holdings and an institution's place on the list. An even stronger correlation with volumes per student indicates that small-enrollment institutions with large library holdings are even more likely to be listed among the top schools (see Table 1). However, a counterintuitive result is obtained when analyzing the correlation of the number of computers available to students and an institution's ranking (see Table 2). Although the correlation is small, its direction is counterintuitive and puzzling. It suggests that an institution suffers a *penalty* in its ranking when the number of computers it provides to students is large.

There are several possible explanations for this. One may be that the data compiled by *U.S. News* for institution-provided computers may be in error. Or there may be ambiguous definitions about what constitutes a computer available for use by students at a university. Some universities may report all campus computers as part of the data, whereas others confine their reporting to formal computer laboratories. Notwithstanding such potential procedural errors on the part of *U.S. News* researchers, some institutions may be much more likely to encourage computer ownership by students themselves, thus obviating the need for provision of computers by the institution. Consider, for example, that it is highly more likely for a matriculating student at MIT to already own a computer than a student at a less technologically oriented institution. Some universities even require students to own a computer. Computer ownership is strongly correlated

TABLE 2  
Correlation of *U.S. News and World Report's*  
Institutional Ranking With Computers

<i>Variable</i>	<i>Rank</i>
Computers	0.110*
Computers/student	0.182**

\*Not statistically significant. \*\* $p = .05$ .

with family income, so a student attending an expensive private school might be more likely to own a computer than a student who attends a community college. No data are offered by the *U.S. News* data about the age, condition of repair, or Internet connectivity of the institution-provided computers. Finally, there may be a “chicken-and-egg” problem confounding the analysis. Some middle-range institutions, looking to move up the *U.S. News* list and fearing their rankings may fall if they don’t implement bold technological solutions, may be finding themselves in the role of technological pioneers, whereas some institutions near the top of the list may pursue a “technology follower” strategy,<sup>2</sup> waiting to see where the trend will go before committing fully. Although the answers to these conjectures are far from clear, the *U.S. News*’s data certainly do not support an unequivocal endorsement of the thesis that computer availability positively affects perceived institutional quality.

Neither do the variables internet availability and e-mail availability provide any illumination about *differential* institutional quality, at least with respect to the *U.S. News*’s listing. Of the top 125 schools that responded to the *U.S. News* survey on this question, *every* school reported providing Internet and e-mail access to all students. Furthermore, for the U.S. Department of Education National Center for Education Statistics’s (National Center for Education Statistics, 2002) listing of 237 U.S. institutions of higher education with enrollment greater than 15000 students, *every single school* had a functioning Web site. Apparently, a school Web site, along with provision of Internet and e-mail for students is considered to be an essential in U.S. higher education today.

## HOW INSTITUTIONAL “BEST OF” LISTINGS COMPARE

*U.S. News* is not the only publication with sufficient audacity to tackle such a controversial subject as prioritizing institutional quality. Other lists, using different measurement criteria, purport to do precisely the same thing. Some examples are Shanghai Jiao Tong University (2003), Gourman (1996), *Webometrics* (<http://www.webometrics.info>), and the *London Times Higher Education Supplement* (<http://www.thes.co.uk>). Each of these “best of” listings uses different criteria for ranking higher education institutions, with very limited overlap in their criteria. The Jiao Tong University listing used five independent variables in its ranking – Nobel laureates, highly cited researchers, articles published in *Nature* and *Science*, articles in Science and Social Science Citation indexes, and academic performance per faculty. The *London Times* used five different criteria: peer review scores, recruiter scores, number of international faculty and students, faculty:student ratios, and citations/faculty in its priority scheme. Gourman used

---

<sup>2</sup>Benefits of pursuing a technology follower strategy are addressed in Christensen (2000).



TABLE 3  
*Webometrics* Ranking of World Universities: Methodology

**Size:** Number of pages is calculated using four engines: Google, Yahoo, MSN and Teoma. For each engine, results are normalized to 1 for the highest value. Then for each domain, maximum and minimum results are excluded and every institution is assigned a rank according to the combined sum.

**Visibility:** The total number of unique external links received (inlinks) by a site can be only confidently obtained from Yahoo and MSN only. For each engine, results are normalized to 1 for the highest value and then combined to generate the rank.

**Rich Files:** After evaluation the "academic" relevance and the volume of different file formats we considering for our purposes the following 'rich files':

EXTENSION	FILE
.pdf	Adobe Acrobat PDF
.ps	Adobe Postscript
.doc	Microsoft Word
.ppt	Microsoft Powerpoint

These data were extracted using Google and merging the results for each file type after normalizing them in the same way as described before. The three ranks were combined according to a formula where each one has a different weight:

$$\text{Webometrics Rank (Position)} = 2 * \text{Rank(Size)} + 4 * \text{Rank(Visibility)} + 1 * \text{Rank(Rich Files)}$$

$$\text{WR} = 2\text{S} + 4\text{V} + \text{R}$$

*Note.* Source: <http://www.webometrics.info/methodology.html>.

18 different and varied criteria, though only 3 of those criteria—library number of volumes, appropriateness of materials to individual disciplines, and accessibility of materials; computer facility sufficient to support current research activities for both faculty and students; and sufficient funding for research equipment and infrastructure—relate either directly or indirectly to the role of Internet and computers in higher education. The *Webometrics's* list is the one most directly tied to the presence or absence of Internet and computer technologies. This ranking is compiled by analyzing institutional Web sites on the basis of size, visibility, and "richness."<sup>3</sup> Explanations of these criteria are shown in Table 3.

Comparison of these various listings using correlation analysis generates an interesting result. Correlation coefficients for the five lists were calculated for the entire number of schools in common on each list, the Top 100 schools, Top 50, Top 20, and Top 10 schools.<sup>4</sup> The results of these correlation calculations are shown in Table 4.

<sup>3</sup>Not to be confused with the notion of "media richness," which is a vibrant media and communications research field of its own. See Daft and Lengel (1986) and Trevino, Lengel, and Daft (1987).

<sup>4</sup>Not all lists had more than 50 schools in common, thus it was impossible to calculate a correlation coefficient for this item.

TABLE 4  
Correlation of Rankings of Worldwide Universities on Five Lists

Lists	Webometrics			Jiao Tong			Times Higher Education						US News		
	Top	Top	Top	Top	Top	Top	Top	Top	Top	Top	Top	Top	Top	Top	Top
	All	100	50	20	10	All	100	50	20	10	All	100	50	20	10
Webometrics	0.472	0.281	0.292	<b>0.485</b>	0.359	0.229	0.319	0.260	0.136	<b>0.560</b>	0.376	0.344	0.322	0.592	<b>0.610</b>
Jiao Tong	0.472	0.281	0.292	<b>0.485</b>	0.359	0.385	0.379	0.324	0.179	<b>0.529</b>	0.516	0.516	0.354	0.563	<b>0.663</b>
Times	0.229	0.319	0.260	0.136	<b>0.560</b>	0.385	0.379	0.324	0.179	<b>0.529</b>	0.381	X	0.368	0.220	<b>0.425</b>
US News	0.376	0.344	0.322	0.592	<b>0.610</b>	0.516	0.516	0.354	0.563	<b>0.663</b>	0.381	X	0.368	0.220	<b>0.425</b>
Gourman	0.422	X	0.422	<b>0.624</b>	- <b>0.359</b>	0.487	X	<b>0.487</b>	0.472	0.432	0.638	X	X	0.707	<b>0.713</b>

Note. Source: Calculated from *U.S. News*, Jiao Tong, Gourman, *Webometrics*, and *Times Higher Education Supplement* listings.

What conclusions, if any, can be drawn from this information? One observation is that the correlation coefficients are almost always higher for the Top 10 or 20 schools than for the Top 50, Top 100, or all schools. This evidence confirms the intuition that there is a general consensus that schools like Harvard, Yale, and Stanford (and, internationally, Cambridge, Oxford, the Swiss Federal Institute of Technology, France's *Ecole Polytechnique*, and Tokyo University) are deserving of high ranking on any list of worldwide universities. However, the apparent consensus about institutional quality tends to disappear after about the first 20 schools.<sup>5</sup> This result offers equivocal evidence, at best, that Internet and computer availability are perceived to strongly affect institutional quality as measured by an institution's ranking on "best of" lists.

### OUTCOME MEASURES AS A GAUGE OF INSTITUTIONAL QUALITY

Because institutional rankings indicate an apparent consensus about the Top 20 or so worldwide universities but indicate increased controversy about institutional quality beyond those institutions, the question becomes, what measures might be used as a reliable indication of institutional quality? Many such discussions of institutional quality devolve to *outcome measures*.

Outcome measures are intended to answer such questions as, "When a student attends a higher education institution, how does the institution know what the student has learned from the experience, and is this learning measurable?" But given the multitude of missions that higher education institutions are called upon to accomplish, is it possible to believe that a consensus might ever be reached on what the outcome of a university education ought to be?

Harvard University president emeritus Derek Bok believes some consensus does exist and that some higher education outcomes are measurable.

[Presently,] applicants to universities have no way of knowing how much they will learn at the college or professional school they are considering, let alone comparing it to how much they might learn at some other institution. . . . Although reliable, universally applicable tests do not exist, and though some educational outcomes cannot be measured at all, tools are already available that can help campuses assess such important competencies as critical thinking, writing, quantitative reasoning and proficiency

---

<sup>5</sup>It is interesting that the Gourman listing and the *Webometrics* listing differ enough on the Top 10 schools to create a *negative* correlation coefficient, the only one in all of the comparisons that is negative. However, this could happen if the same schools were listed in the Top 10 but their orders were reversed—for example if Harvard is listed as number 1 on one list and number 10 on the other list. Thus, the *absolute value* of the correlation gives more information about the consensus of the Top 10 than their specific positions.

in foreign languages. These measures may not be perfect, but they are a big improvement over knowing little or nothing about student progress. Many institutions use such instruments already. Others participate in national surveys to determine where they stand in making use of the most effective methods of teaching and learning. ([www.forbes.com/2006/04/15/derek-bok-university\\_cx\\_db\\_06slate\\_0418bok.html](http://www.forbes.com/2006/04/15/derek-bok-university_cx_db_06slate_0418bok.html))

Also embedded within Bok's comments can be found a reasonable definition for a outcome quality measure for higher education, namely, the ability to teach students to "write better, speak more eloquently, think more rigorously, or reason quantitatively more proficiently." Bok seems to think that these kinds of outcomes can be measured and that technology can assist not only in helping students attain these skills and knowledge but also in their measurement.

## INTERNET AND COMPUTER TECHNOLOGY GROWTH WORLDWIDE

Bok, as just quoted, and many others (see, e.g., Castro, 2000; Duderstadt, Atkins, & Van Houweling, 2002; Phipps, 2004) agree that Internet and computer technologies can enhance higher education outcomes. This was also confirmed through interviews with numerous higher education faculty and administrators at five institutions in the Washington, DC/northern Virginia area ( $n = 17$ ). Therefore, one way to analyze the likely future of the higher education gap is to analyze the differential penetration of Internet and computer technology at institutions in both low- to middle- and high-income countries—and to determine what direction is the trend going—toward a wider or narrower gap.

When the analysis is done at a national level, the emerging trend is that global Internet use is not nearly as strongly dominated by users in high-income countries as it was in the early days of the Internet. When Internet usage data for the United States and for high-income countries is plotted in the same chart with *all* users, it is apparent that the gap is closing, albeit slowly. Figure 1 illustrates that the United States had the lion's share of Internet use in the early days of the Internet (mid- to late 1990s) but that the United States' percentage of the world's total has been declining ever since.

The same trend appears when all high-income countries (the United States, Canada, western Europe, Japan, Korea, Singapore, Taiwan, Australia, and New Zealand) are added together on the same graph and plotted with world usage. Though the high-income countries dominated well into the late 1990s, it is apparent now that their percentage of the whole is declining (see Figure 2).

This implies that Internet use growth is accelerating in low- to middle-income countries and leveling off in high-income countries. In fact, a list (see Table 5) of the Top 25 countries in per-capita growth in Internet use from 2000 to 2004 is dominated by low- to middle-income countries.



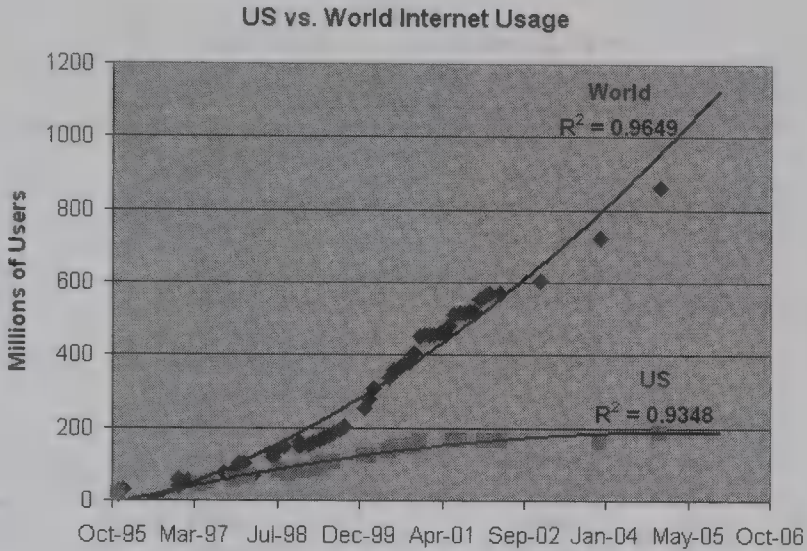


FIGURE 1 Internet usage in the United States versus the World—1995–2006.

But this analysis is a bit simplistic. What emerges from a deeper investigation is that low- to middle-income countries exhibit a diversity of Internet usage growth patterns, with likely plateaus at different levels, ranging from 1% penetration levels to greater than 50% penetration over the long term (see Figure 3). This follows an

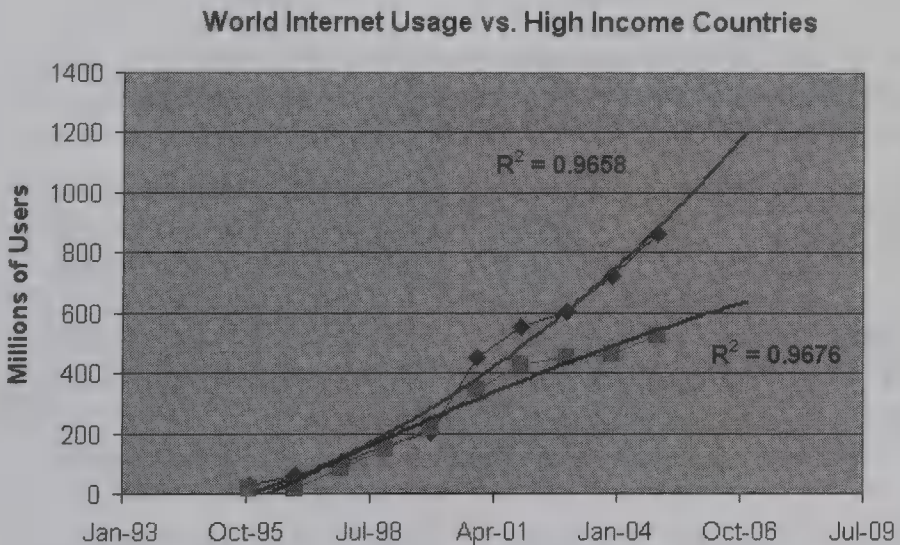


FIGURE 2 Internet usage in high-income countries versus the World—1995–2006.

TABLE 5  
Top 25 Countries in Internet User Growth:  
2000–2004

No.	Country	Users' Growth Rate 2000–2004
1	D.R. Congo	9900%
2	Haiti	8233%
3	Somalia	7400%
4	Congo	7100%
5	Sudan	3700%
6	Guyana	3525%
7	Azerbaijan	3300%
8	Vietnam	2835%
9	Iran	2100%
10	Albania	2043%
11	Martinique	2040%
12	Libya	1950%
13	Syria	1933%
14	French Guiana	1800%
15	St Lucia	1733%
16	Morocco	1650%
17	Algeria	1590%
18	Zimbabwe	1540%
19	Guam	1480%
20	Barbados	1400%
21	Pakistan	1394%
22	Dominican Republic	1355%
23	Belarus	1267%
24	Jamaica	1234%
25	Bhutan	1233%

*Note.* Source: International Telecommunication Union (<http://www.itu.int>).

approximate 4- to 7-year lag due to early rapid growth in high-income countries.<sup>6</sup> Internet usage penetration, in addition to being strongly correlated with national income ( $r^2 = .885$ ,  $n = 145$ ,  $p = .01$ ), exhibits equally strong correlation with other national-level variables: teledensity<sup>7</sup> ( $r = .909$ ,  $n = 128$ ,  $p = .01$ ), tertiary education enrollment percentage (.826,  $n = 78$ ,  $p = .01$ ), and corruption<sup>8</sup> (.884,  $n = 100$ ,  $p = .01$ ). A regression analysis showed that these latter three independent variables were sufficient to predict Internet use levels within 85% of the variation of the dependent variable, without introducing income into the equation.

<sup>6</sup>Reasons for differential long-term adoption levels are discussed in depth in Capshaw (2007).

<sup>7</sup>Teledensity is defined as the number of phone lines per capita in a country.

<sup>8</sup>As reported by Transparency International's (2004) Corruption Perception Index.

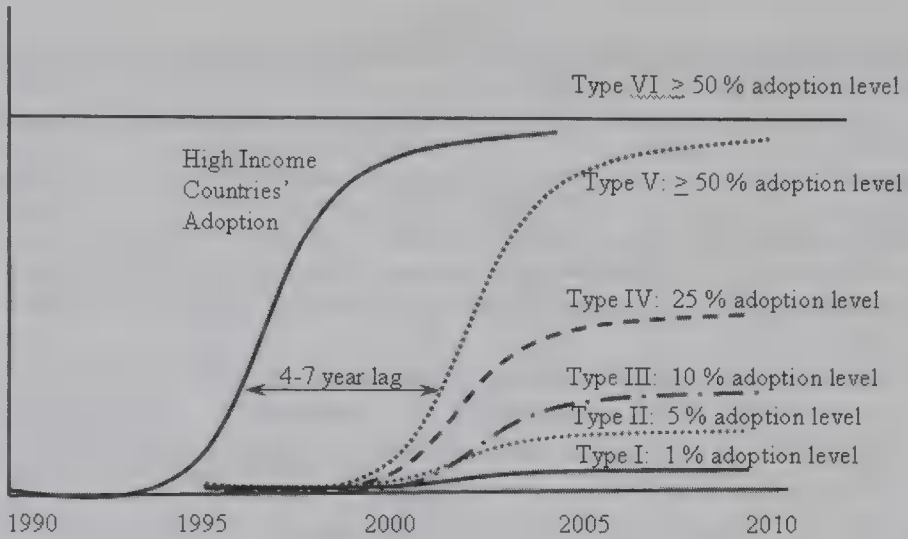


FIGURE 3 Long-term plateau levels for Internet usage in low- to middle-income countries.

One conclusion from the foregoing is that a country's income level is not necessarily a determinant of its ability to become connected to the Internet. Other national-level variables, though they may be connected to income, play at least as important a role in that determination.

## INTERNET TECHNOLOGY AT UNIVERSITIES WORLDWIDE

Turning now to look at higher education institutions in low- to middle-income countries, one may ask, "What are the indicators of the use of Internet and computer technology at the institutional level, and what is the progress of this technology at these institutions in comparison to institutions in the United States?"

One face that a higher education institution presents to the community that provides a strong indication of its technological sophistication is the institution's Web site. It was noted previously that there are 237 higher education institutions in the United States with an enrollment greater than 15,000 students—and every one of these institutions has an active Web site. It would therefore be possible to understand something about these institutions by looking in detail at these Web sites. A random sample ( $n = 147$ ,  $p = .05$ ) of these sites was analyzed and rated on the following five dimensions:

1. Does the Web site enable electronic access to research tools: the university library or electronic databases/journals?

TABLE 6  
Random Sample of U.S. Institutional Web Sites Rated on Five Dimensions of Content

Dimension of Information	1	2	3	4	5
	Library	Virtual Education	Administrative Information	E-mail	Course Information
% of Web sites having information	100%	98.6%	100%	86.4%	100%

Note. Source: 147 U.S. institutions of higher education with greater than a 15,000 enrollment.

- 2. Does the Web site offer online virtual learning or a Course Management portal such as *Blackboard*, *WebCT*, or *Prometheus*<sup>9</sup>?
- 3. Does the Web site provide basic administrative information about the institution (where it is located, how to apply)?
- 4. Does the Web site enable student e-mail access?
- 5. Does the Web site provide detailed information about academic departments and course offerings at the institution?

For the random sample of U.S. institutions, all Web sites were available when accessed via the Internet, and the percentage of each dimension of information is shown in Table 6.

The results of this sample are striking—every single institutional Web site in the sample had basic administrative information about the institution, detailed department and course information, and electronic access to the library. Almost all had a distance education portal or a Course Management system. Most provided institutional email access for the students,<sup>10</sup> and those institutions that did not offer such institutional e-mail access had apparently decided that there were many free e-mail options available for students and therefore it was unnecessary for the institution to provide this service.

In addition to these five dimensions, there were a number of other services and information sources provided through institutional Web sites, such as access to grades and transcripts; tuition payments and financial aid; online writing and research help; personal Web pages for faculty and students; staff, faculty, and student directories; housing, parking, and university police information; and even online advising, tutoring, and video tours of the campus.

A similar sample was taken from a list of approximately 2,500 low- to medium-income country higher education institutions listed by the *Universities Worldwide*

<sup>9</sup>Although these “Course Management Systems” potentially provide a portal for online distance education, they are also often used as a supplement to classroom-based courses at many U.S. institutions.

<sup>10</sup>Usually an e-mail address with an institutional tag—such as [ClarkCapshaw@vanderbilt.edu](mailto:ClarkCapshaw@vanderbilt.edu) and the ability to access this account through the institutional Web site.



TABLE 7  
Random Sample of Low- to Middle-Income Country Institutional Web Sites Rated on Five Dimensions of Content

<i>Dimension of Information</i>	<i>1 Library</i>	<i>2 Virtual Education</i>	<i>3 Administrative Information</i>	<i>4 E-mail</i>	<i>5 Course Information</i>	<i>Sites Not Available</i>
Latin America and Caribbean	61.9%	46.7%	99.0%	58.1%	84.8%	5
Asia	46.0%	24.6%	100.0%	55.6%	79.4%	18
Middle East	42.9%	28.6%	100.0%	52.4%	95.2%	2
Africa	38.9%	25.0%	100.0%	50.0%	66.7%	1
All Countries	50.5%	32.9%	99.7%	55.7%	81.0%	26
PREVALENCE RANK	4	5	1	3	2	

Web site (<http://univ.cc/world.php>). Seventy percent of these Web sites were active when accessed in November and December 2005. From these 1,733 Web sites, a random sample of 314 was chosen to ensure a .05 significance level. The 314 sites were analyzed for the same five dimensions of information as in the sample of U.S. institutions, and Table 6 shows the results. For these Web sites, even though all of them were available when accessed in December 2005, several were not available in January or February 2006 when tested for content; this occurrence and other evidence leads to the conclusion that low- to middle-income countries are sometimes only intermittently available. In all, 288 Web sites from the sample of 314 were tested for content (91.7% of the sample).

At the bottom of Table 7 is another ranking—the relative percentages of the types of information available on the Web sites. Note that almost all institutional Web sites give administrative information, many give detailed department or course information, and substantially fewer give e-mail and electronic library access. Virtual education and Course Management systems are the least prevalent at institutional Web sites in the developing world.

## INTERPRETATION AND EXTENSION OF THE RESULTS OF THE WEB SITE ANALYSIS

The model in Figure 4 was developed based on the study of institutional Web sites in the United States and in developing countries,<sup>11</sup> and the results of interviews

<sup>11</sup>The process of such Web site development may follow a multitude of models, but one such model that is likely to be prevalent today in some low- to middle-income country institutions is the following: a graduate student or even an undergraduate student who is technologically adept plays a strong role in the initial development of the school's Web site. He or she is limited in this effort in several respects: by the information provided by the institution for inclusion in the Web site, by the hardware and software

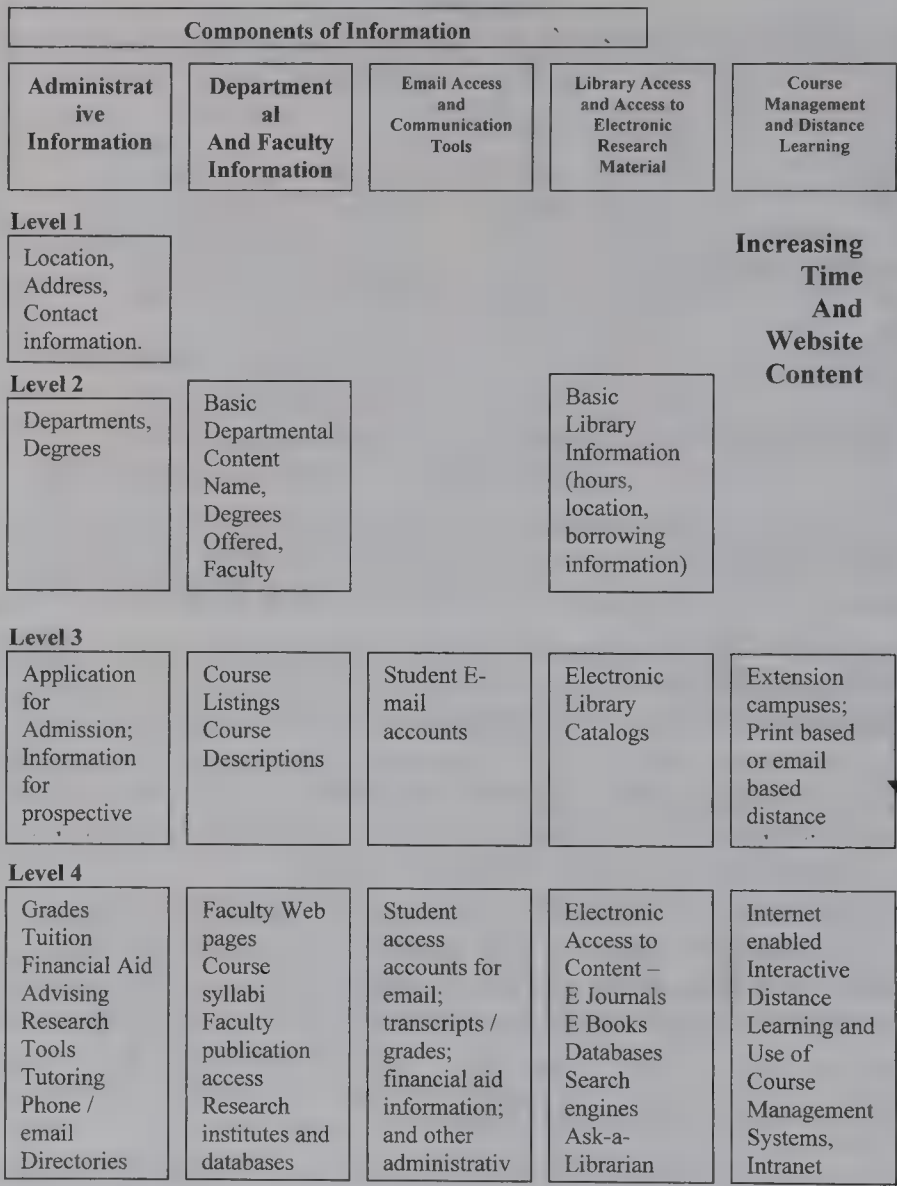


FIGURE 4 The evolution of information content of an institutional Web site.

with higher education administrators and faculty at several U.S. institutions who have had firsthand experience with the development of such tools over time. Most institutions begin the development of the institutional Web site by providing very basic administrative information—the location/address of the school and

that he or she uses to initially develop the Web site, by the bandwidth limitations due to the country or institution’s infrastructure, by his or her own knowledge and ability, and by his or her eventual departure from the institution through graduation or transfer, often leaving no one to maintain the Web site.

admission procedures being the information that is deemed to be the most essential. Therefore, it is highly likely that if a Web site exists it contains administrative information of this type at a minimum. In Figure 4, this represents the first level of Web site development.

As institutions continue to develop their Web sites, often the next step is to introduce more detailed departmental or course information, along with a parallel effort to develop more sophisticated administrative information: online applications, messages from institutional officials, and information for prospective students. Along with this effort may be an initial attempt to connect to the institution's library—although more often than not, these connections offer only rudimentary information about the library: location, hours, and occasionally electronic card catalogs of library holdings. In this instance, a student would still have to physically visit the library to gain access to most of the resources. This level of Web site sophistication represents the second level from the top in the figure. Level 3 in the diagram represents yet another iteration of sophistication—some institutions provide online admissions applications, detailed course listings and catalogs, student e-mail accounts, and more library access. As indicated by the data from the Web site analysis, many of the low- to middle-income country institutional Web sites are reaching this level of sophistication, but all U.S. institutions have already passed this level of sophistication. All U.S. institutions that were analyzed are already at Level 4 in the diagram—offering a variety of administrative information and services, detailed departmental and course information, multiple student services often available from a sign-in service, multiple electronic journal and database access through the library, and Course Management systems and sophisticated means to deliver course content electronically at a distance for e-learning. But the evolution of low- to middle-income country institutional Web sites is seen to be developing along the same lines, and quite a few institutions are already offering sophisticated information and services through their Web sites that rival that of the U.S. institutions. In short, it is the same phenomenon that was noted with national Internet connectivity—these countries lagged behind at the start but are catching up. Simultaneously, they are learning lessons from the experience of institutions in high-income countries. The question is, will their overall long-term progress be limited, and if so, to what extent?

## CONCLUSIONS

The central question of this research, whether use of Internet and computer technologies will enable higher education institutions in low- to middle-income countries to close the quality gaps with higher education institutions in high-income countries, is not one that can be answered without some degree of equivocation, because so much depends on decisions that must be made to maximize the potential

of these technologies. But, it has been shown that low- to middle-income countries are already closing the gaps in Internet connectivity at the national level and that higher education institutions in these countries have already begun to develop some sophistication in the use of these technologies. The ultimate impact on quality is undetermined, and much controversy still attends the definition and measurement of institutional quality. Notwithstanding these caveats, it is increasingly likely that such technologies will, at a minimum, extend *access* to higher education to a wider range of students, and through the ability of the Internet and computer technologies to provide access to greater amounts of information, it will enable low- to middle-income country institutions to pass this greater content knowledge along to their students, whether through traditional rote methods or by transitioning to the more critical thinking, constructivist model now used in high-income countries.

## REFERENCES

- Capshaw, N. C. (2007). *The Quality of higher education: Internet and computer technologies—exacerbating or lessening differences across countries? An analysis at three levels: National, institutional, and classroom*. Unpublished dissertation, Vanderbilt University, Nashville, TN.
- Castro, C. D. M. (2000, January/February). Technology and institutional change: Why some educational institutions use technology and others don't. *TechKnowLogia*, pp. 14–15.
- Christensen, C. (2000). *The innovator's dilemma: When new technologies cause great firms to fail*. New York: Harper Business.
- Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness, and structural design. *Management Science*, 32, 554–571.
- Duderstadt, J. J., Atkins, D. E., & Van Houweling, D. (2002). *Higher education in the digital age: Technology issues and strategies for American colleges and universities*. Westport, CT: Praeger.
- Gourman, J. (1996). *The Gourman report: A rating of undergraduate programs in American and international universities (9th ed.)*. Los Angeles: National Education Standards.
- Jones, S. (2002). *The Internet goes to college: How students are living in the future with today's technology*. Washington, DC: Pew Internet & American Life Project.
- National Center for Education Statistics. (2002). *Digest of education statistics, 2002*. Retrieved June 26, 2004, from <http://nces.ed.gov/programs/digest/d02/tables/dt413.asp>
- Phipps, R. A. (2004). *How does technology affect access in postsecondary education? What do we really know?* Washington, DC: National Postsecondary Education Cooperative Working Group on Access-Technology.
- Shanghai Jiao Tong University. (2003). *Academic ranking of world universities*. Retrieved August 2, 2004, from <http://ed.sjtu.edu.cn/ranking.htm>
- Transparency International. (2004). *Global corruption report, 2004*. Berlin, Germany: Author.
- Trevino, L. K., Lengel, R. H., & Daft, R. L. (1987). Media symbolism, media richness, and media choice in organizations. *Communication Research*, 14, 553–574.



# Compulsion, Craft, or Commodity? Education Services Trade in the Larger Context

Brandyn L. Payne

*Healthways, Incorporated Nashville, TN*

The role of education in fostering economic growth and social development is universally recognized. Although history places the provision of education firmly within national control, countries increasingly search outside national borders for alternative distribution frameworks. Tellingly, the World Trade Organization recently included education as service trade sector in the General Agreement for Trade in Services (GATS) negotiations. Such activity increases debate about control as countries struggle to create policies that balance nationalism with economic responsiveness. This study employed multivariate data to question whether trade openness in 162 countries was associated with openness to trade in education, and whether countries' commitments to lower barriers to education trade paralleled the strength of their commitments to lower barriers to all trade.

Among the findings were the following: (a) On average, countries with education commitments experienced slightly higher levels of general trade openness than those without education commitments; (b) in lower-middle-income countries, education trade openness and general trade openness were positively related; and (c) when controlling for education, population, geography, and income, lower levels of education trade barriers were the single best predictor of countries' having made education commitments under GATS. A model for systemic improvement in education trade policymaking is also presented.

The critical role of education in fostering economic growth and social development is universally recognized. However, cultural and ethical concerns continue to inspire education debate. Although historical precedent places the provision of education firmly within national control, heightened access and efficiency

requirements increasingly drive countries to search outside national borders for higher quality, equitable access and improved distribution frameworks.

Such searches shape current domestic education policy, as evidenced by dramatic growth in the private provision of education products, services, and programs. Without exception, industrialized democracies have elected to contract with nongovernmental suppliers for textbooks, educational software, testing, administrative activities, and countless other products and services.

These supply relationships often span international borders, constituting a growing element of international trade. In 2004, the United States alone exported over \$13.5 billion in education services, an 11% increase over 2003 (U.S. International Trade Commission, 2006). As further evidence of this growth, the World Trade Organization (WTO) recently included education as sector of service trade within the General Agreement for Trade in Services (GATS) negotiations (WTO, 2001). Such activities increase the intensity of debate over who controls a nation's education (Heyneman, 2001, 2003; Jarvis, 2000; Larsen, Morris, & Martin, 2002; Lenn, 2000; Sauve, 2002; WTO, 2001), as nations struggle to build education policies that balance nationalism with economic responsiveness.

Critics suggest that trade in education abrogates a nation's right to provide for its own citizens (Larsen et al., 2002, and others). Others suggest that wealthy or well-positioned nations will dominate trade, threatening the existence of local cultures, languages, and learning priorities (Altbach, 2001, 2002, 2003; Hill, 2001; Naidoo, 2007; Nyborg, 2002; Van Den Wende, 2001).

But is education trade really all that different from consulting, telecommunications, or information technology trade? Did countries that have made education services commitments under GATS consider education's unique value when making commitments, or were they more likely to propose and support policies that mirrored their general trade agendas? Or put differently, is the widely proposed view that education cannot be considered as a service to be traded even valid?

To discover what factors are associated with a nation's trade policy in education, it is first necessary to ask, What is the nature of the relationship between countries' openness in education trade and their position on general trade issues? More specifically, Is education trade openness a component of larger trade openness, and what characteristics are associated with countries that have already made education commitments?

Why are these questions important? The subject of trade in higher education services often inspires debate and confusion among decision makers, particularly as need for access continues to grow. In recent years, several researchers (see Knight, 2002a, 2002b; Larsen et al., 2002; Lenn, 1999; Lenn & Miller, 2000; Sauve, 2002; Van den Wende, 2001, for examples) and agencies (see American Council on Education, 2004; Organisation for Economic Cooperation and Development [OECD], 2002; U.S. Department of Commerce, 1998, 2000; WTO, 1999a, 1999b, for examples) have published work on GATS and liberalized trade

and its impact on education. However, much of this work is declarative, designed to inform researchers, policymakers, and the public about the provisions of the agreement, the current barriers to trade, and potential benefits or drawbacks. Increasingly, researchers are calling for more rigorous analysis of the potential risks and opportunities of increased education trade (Knight, 2002a, 2003; Larsen et al., 2002; Nguyen-Hong & Wells, 2000; Sidhu, 2007) as well as stronger measures and collection instruments (Ascher, 2001; Asia-Pacific Economic Cooperation, 2001; Kemp, 2001; Knight, 2003; OECD, 2002; Sauve, 2002; WTO Secretariat, 2001, and others).

This study seeks to add to the limited body of current research while using the ongoing WTO–GATS negotiations as a reference point for discussion of trade barriers, openness, and policymaking. To strengthen the analysis, this study also draws heavily from other disciplines, including economics (Pritchett, 1994; Rose, 2002) government, and education (Kemp, 2001; Larsen et al., 2002; McGuire & Schuele, 2000).

### “OPENNESS” AS INDICATOR

Central to the notion of increased mobility is the idea of “openness” in a country’s trade policy. A common measure of trade openness is the ratio of imports and exports divided by aggregate Gross Domestic Product (trade/GDP) for a particular moment in time, defined by Pritchett (1994) and others as the *trade intensity* of a particular economy. Economists and educational researchers alike have struggled to measure the effects of trade policy on openness and growth (Dollar, 1992; Dollar & Kraay, 2001; Edwards, 1997; Greenaway, Morgan, & Wright, 2002; Sachs & Warner, 1995). In their 2001 study, “Trade, Growth, and Poverty,” Dollar and Kraay asked, “What can we expect to happen when developing countries liberalize trade and participate more in the global trading system?” They found that increased trade openness led to faster economic growth and improved standards of living for millions of the world’s poor.

In a 1994 article, Pritchett used 16 potential measures to assess outward orientation for lesser developed countries, including policy incidence, average tariff levels, structure-adjusted trade intensity, Leamer’s Openness Index, and trade and price distortion. He found none of these measures to be significantly useful for measuring openness for the 168 countries present in the Penn World Table (PWT).

In 2002, Andrew Rose used Pritchett’s individual variables, along with a trade/GDP measure, as openness indicators in a study analyzing links between trade openness and WTO membership. Rose concluded that little evidence existed that WTO member countries had more liberal trade patterns than nonmember countries (see Figure 1).

$$Open = \frac{X + M}{D_Y + X - M}$$

$X$  = Value of all exports  
 $M$  = Value of all imports  
 $D_Y$  = Total of all domestic consumption and investment, public and private; GDP

FIGURE 1 Formula for openness (Pritchett, 1994).

Any study of trade growth must consider that some researchers guard against using the trade/GDP ratio and/or the concept of “openness” as a basis for classifying countries’ trade policies as open or closed to outside providers. Birdsall and Hamoudi (2002) argued that for countries that are highly dependent on commodities for their export revenue, the trade/GDP ration overstates the importance of trade policy in economic growth. Although this may be the case, the acceptability of openness within the education trade community, including its use in recent studies of education trade (see, e.g., Kemp, 2001; Nguyen-Hong & Wells, 2000), renders it appropriate for this analysis.

METHODOLOGY AND RATIONALE

This study uses descriptive and inferential statistics to test the hypothesis that education openness is not a function of overall trade openness. These measures are consistent with recent literature analyzing the relationship between trade openness and a variety of factors (Edwards, 1997; Greenaway et al., 2002; McGuire & Schuele, 2000; Rose, 2002; Sachs & Warner, 1992).

One would assume that if a positive relationship exists between education trade openness and general trade openness, education trade currently functions as a component of a country’s larger trade context (see Figure 2). If no relationship, or a negative one, is found, one may conclude that education is operating in a different trade context from overall trade efforts for the countries in this sample.

The sample for this analysis is composed of all countries included in Rose’s 2004 study correlating openness with World Trade Organization membership

Sample	A ←	→ B
All countries (n=162)	Trade Openness (Rose, 2003)	GATS Education Commitments - Y/N (WTO, 1999)

FIGURE 2 Is education openness a function of general trade openness?



$$r_{pb} = \frac{M_1 - M_0}{s_y} \sqrt{pq}$$

Where:

$r$  is the point-biserial correlation coefficient

$M_p$  is the mean general trade openness for countries with education commitments

$M_q$  is the mean general trade openness for countries without education commitments

$s$  is the standard deviation for openness values

$p$  is the fraction of countries with education commitments

$q$  is the fraction of countries without education commitments.

FIGURE 3 Formula for point-biserial correlation (Chen & Popovich, 2002).

( $n = 162$ , taken from the PWT, version 6.1). The PWT database was used to capture trade openness statistics over time by selecting five instances over the past 20 years, beginning with 1980 and ending with 2000. The point-biserial correlation coefficient, a particular type of correlation statistic used to estimate the relationship between a continuous variable (overall trade openness) and a naturally dichotomous variable (in this case, the presence or absence of education trade commitments under GATS), was used to conduct this correlation analysis. Results from these procedures are described in the Findings section of this article.

To answer the second research question (What characteristics are associated with countries that have made education commitments?), a set of regression techniques were used to compare the dependent variable of commitments to education services trade against a variety of explanatory variables, including presence of barriers to education trade, foreign enrollment, and general trade openness while controlling for geographic and economic differences between countries. This investigation is consistent with recent, if limited, studies analyzing the impact of educational services trade (Kemp, 2001; Larsen et al., 2002; Nguyen-Hong & Wells, 2000; see Figure 3).

In the case of the second question, one would hypothesize that explanatory variables have differing levels of effect on countries' probability of having made commitments to education trade. In reviewing the literature, it has been suggested that education barriers are an important consideration in countries' willingness to make commitments to education trade (Kemp, 2001; Nguyen-Hong & Wells, 2000). However, it may be the case that other factors, including foreign enrollment, the subsectors in which a country chooses to focus commitment, or even a country's overall trade volumes may have a greater impact on the outcome. Similarly, variables that are not currently collected at a discrete level—such as subsector with the greatest export movement, education services import and export revenue, and private investment in education services—may influence countries' likelihood of having made education commitments.

### *Data Availability*

This study uses secondary data available through research literature and the Internet. The specific data set used to report overall trade openness values is the PWT, version 6.1, maintained by the Center for International Comparison at the University of Pennsylvania. The PWT reports purchasing power parity, international pricing statistics, and other basic economic indicators for 168 countries from 1950 to 2000. Data from the PWT are used by the Organization for Economic Cooperation and Development (OECD), the European Union, UNESCO, the World Bank, and other global organizations to report economic data for domestic and international trending and tracking purposes (Heston, Summers, & Aten, 2002). In an attempt to describe the recent education trade landscape, the 2000 PWT data collection was used to answer both research questions.

Openness data for a handful of countries were not available through the 2000 PWT sample. For these countries, publicly available UNESCO and OECD data were used to generate openness measures for 2000 (these substitutions are noted in the technical notes listed in the appendix). In addition, data specifically related to higher education when forced to make a choice about which subsectors on which to report. The rationale for this, consistent with the rest of this study, is that higher education represents the largest and most aggressive subsector of the education services market.

Data for the dependent and explanatory variables used in the regression analysis were also collected through publicly available sources, including databases maintained by the WTO, UNESCO, OECD, and the World Bank. Individual variables are operationalized in the following section with their original sources and any alternative collection methods noted.

### *Operationalization of Variables*

Overall, one independent variable and five dependent variables were used in this analysis. Unless noted, 2000 is used as the baseline year for all observations. Variables restricted to a particular subsector of education trade reflect higher education statistics.

*[WOPEN]: Overall Trade Openness (2000).* WOPEN is a continuous, independent measure of individual countries' overall trade openness. This is a commonly accepted measure of an individual country's "openness" to outside goods and services as well as the impact of this cross-national trade on overall economic health. WOPEN is used as the basis for correlation in this analysis. It is identical to the PWT 6.1 OPENC measure. In the case of countries for which PWT 6.1 data for 2000 were not available, a proxy measure was substituted for WOPEN.

[COMMYN]: *Presence of education commitments (2000)*. COMMYN is a dichotomous, dependent variable representing the presence or absence of education trade commitments under GATS, where 0 is *no commitment* and 1 is *commitment*. For consistency with WOPEN, commitments are reflected as current for 2000–2001. Verification of these commitments was taken from the WTO Services database (<http://www.wto.org>), in which countries' overall service commitments under the Doha Round are represented in matrix form.

[EDBAR]: *Presence of education service trade barriers (2000)*. EDBAR is an independent variable used to quantify the distribution of a particular country's current barriers to education services trade. Its calculation is based on work by Hoekman (1995), McGuire and Schuele (2000), and Kemp (2001). For this analysis, barriers are analyzed specifically for the higher education subsector (see Knight, 2002b, 2003).

Barriers are weighted based on the country's level of commitment to liberalization, using a frequency index developed by Hoekman (1995) and used previously by Kemp (2001). The index is based on GATS commitment schedules and follows a three-value scoring system: a full commitment to liberalize trade is assigned a score of 0, a partial commitment is assigned a value of 0.5, and an unbound commitment is given a value of 1.

Possible rankings range from 0 to 8, with 8 representing the highest presence of barriers to the free import and export of education services. Values represented by the countries in this analysis range from 0 (Congo RP, Lesotho, Sierra Leone, and Slovenia) to 8 (countries with no commitments under GATS). For the three countries for which national-level data were not available, recent publications by the WTO and WTO member nations were used to approximate values for countries in which barriers were thought to be present. In addition, the barrier scores in this analysis were transformed for consistency in interpretation into an inverse scale based on the total possible number of barriers, represented as [8-EDBAR].

[EDCOMW]: *Weighted value of education commitments (2000)*. In contrast to the variable EDCOM, EDCOMW ranks the distribution of a country's trade commitments based on their subsector. This scheme, created by Kemp (2001) to better illustrate the importance of higher education commitments to the overall education services trade debate uses an interval scale of .00 to 1 to quantify the level of commitment. Although primary, secondary, adult, and other education subsectors are assigned a value of .15, higher education receives a value of .4 to denote its position as the most traded sector (remaining sectors are measured at .15) (2001).

Values for countries included in Sample 2 range from 0 (e.g., Estonia) to a perfect 1 (Czech and Slovak Republics, Lesotho, and Sierra Leone).

*[FORENR]: Foreign enrollment as a percentage of overall enrollment (2000).* This continuous dependent variable attempts to quantify education services trade by using foreign enrollment as percentage of overall enrollment in tertiary education. The decision to use this variable as a proxy for overall education trade volume by country was based on a study done by Larsen et al. (2002). In that study, WTO and OECD data were used to approximate education trade as a percentage of overall trade value for OECD countries in 2000. Although the results provided a broader analysis of the overall import and export of education services, education trade data were only available for 11 countries, making any generalization to the larger global community extremely difficult. In contrast, data on foreign enrollment in higher or tertiary education are available for a greater sample of countries, making it a better fit for the research questions pursued in this analysis.

In addition to these variables, eight independent variables were used in the probit regression to control for between-country differences in geography, population, education, and income (see Table 1). These controls are similar to those used by Rose (2002) and others (Kemp, 2001; Nguyen-Hong & Wells, 2000) to mitigate demographic and economic differences between countries that could account for invalid effects. All control variables were pulled from the World Development Indicators database for 2000, and all are used in their original form in this analysis.

### *Educational Openness Index*

No single measure exists to quantify the volume or impact of education services trade for a particular country. Although attempts have been made to quantify the

TABLE 1  
Bivariate Data Summary (Correlating Openness With Education Commitments)

	<i>M</i>	<i>SD</i>	<i>r<sub>bpi</sub></i>
All countries <sup>a</sup>	85.966	43.302	.132
High income <sup>b</sup>	102.807	56.636	-.256
Upper-middle income <sup>c</sup>	113.622	41.787	.148
Lower-middle income <sup>d</sup>	70.280	29.925	.310**
Low income <sup>e</sup>	71.042	28.795	.150

<sup>a</sup>*N* = 162. <sup>b</sup>*n* = 37. <sup>c</sup>*n* = 30. <sup>d</sup>*n* = 46. <sup>e</sup>*n* = 49.

\*\**p* = .05 level (two-tailed).



$$Z_{\text{edcom}} = b_0 + bX_{\text{edbar}} + bX_{\text{wopen}} + bX_{\text{forenr}} + bX_{\text{geo1}} + bX_{\text{geo2}} + bX_{\text{pop1}} + bX_{\text{pop2}} + bX_{\text{ed1}} + bX_{\text{ed2}} + bX_{\text{incom1}} + bX_{\text{incom2}}$$

FIGURE 4 Model for probit regression (Pampel, 2000).

value of overall services trade (Konan & Maskus, 2004; OECD, 2002; and others), the lack of specific education data for most countries makes drawing conclusions difficult. In an effort to focus attention on quantifying education trade (Kemp, 2001; Larsen et al., 2002; Nguyen-Hong & Wells, 2000), the EDBAR, EDCOMW, and FORENR variables were transformed into an index designed to judge countries' relative "openness" as related to the cross-border movement of education services (see Figure 4). After reviewing existing literature, it was determined that such an index could reasonably be constructed from a variety of measures used in recent research (Center for Quality Assurance in International Education, 1999; Knight, 2002b; Larsen et al., 2002; McGuire & Schuele, 2000). Two recent indexes of note are the aforementioned trade restrictiveness index implemented by Kemp (2001) and the set of *trade restrictiveness indexes* constructed by a team of researchers from Australia's Productivity Commission, the University of Adelaide, and the Australian National University (Nguyen-Hong & Wells, 2000). Although this Index was created with the intent that a multivariate model would provide stronger predictive ability than recent, univariate research studies, the lack of available data rendered it virtually useless for the purposes of this analysis. However, it bears mention here as a possible method for strengthening analyses around predictors of education trade, particularly as data quality and quantity increases over time.

### Procedures

Data were collected from the aforementioned publications and online databases during the summer and fall of 2004. Individual countries were identified in the data set by country name and ISO classification code.

Two statistical techniques of note were used in this analysis. The first was a point-biserial correlation, a Pearson product-moment correlation designed to correlate a continuous variable with a dichotomous variable (Brown, 1996). Like the Pearson  $r$ , the  $rpbi$  can range from 0 to +1.00 if the two scales are related positively and from 0 to -1.00 if the two scales are related negatively (or stated differently, in opposite directions). The higher the value of  $rpbi$  (positive or negative), the stronger the relationship between the two variables. The point-biserial correlation is used in this analysis to analyze the relationship between countries' general trade openness (WOPEN) and the presence or absence of education services commitments under GATS (COMMYN; see Figure 5). The traditional Pearson  $r$  was used for additional correlation analyses between quantitative variables.

$$EOI = [1/3 (P_1 + P_2 + P_3)]$$

Where:

$P_1$  = Presence of barriers to education trade [EDBAR]

$P_2$  = Weighted distribution of education services commitments under GATS [EDCOMW]

$P_3$  = Foreign tertiary enrollment as a percentage of overall enrollment [FORENR]

FIGURE 5 Educational openness index (based on work by Kemp, 2001, and Nguyen-Hong & Wells, 2003).

A probit regression model was also employed in this analysis (see Figure 5). Probit coefficients correspond to the  $b$  coefficients in regression or logit coefficients in logistic regression. To interpret the effects of probit, one transforms the coefficients based on the standard normal curve and expresses the results in terms of marginal effects on the likelihood of the probability of a specified value of  $X$  (Pampel, 2000). This difference is called the *elasticity* of the probability of the dependent variable ( $Y$ ) in respect to the independent variable, when all variables are held at their sample means. Elasticity is the effect of a unit increase in the independent variable on the probability that the dependent = 1, when all other independents are held constant at their mean values (Pampel, 2000).

### *Missing Data*

Because of the lack of specific data on education trade collected across all countries, several measures were employed to address missing data in this analysis. Procedures for treatment of missing data are detailed in the larger paper (Payne, 2005).

## FINDINGS

### *Is Education Trade Openness a Component of Larger Trade Openness?*

General trade data comparing countries with education commitments versus those without commitments under GATS is summarized for 1980, 1985, 1990, 1996, and 2000 in Figure 6. Although overall trade openness has trended upward over the past 20 years, countries with education services commitments ( $M = 71.0$ ) experienced significantly greater general trade openness than did countries with no education services commitments under GATS ( $M = 48.0$ ). Results indicated a significant difference in overall mean trade openness,  $t(54.7) = 5.43, p = .001$ .

The point-biserial correlation between general trade openness and openness to educational services trade resulted in an insignificant result,  $r_{\text{pbi}}(160) = .132$ . In an effort to compare results irrespective of national economic characteristics,

$$Pr(y=1|x) = \Phi(xb)$$

Where:

$\Phi$  is the standard cumulative normal probability distribution

$xb$  is the probit score or index

FIGURE 6 Formula for probit regression (Pampel, 2000).

additional correlations were run for groups of countries divided by World Bank income classifications (high-income OECD and non-OECD, middle-upper income, middle-lower income, and low income; World Bank, 2004). It was found that for lower-middle-income countries, a significant positive relationship exists between general trade openness [WOPEN] and the presence of education trade commitments [COMMYN]. That is, only in lower-middle-income countries such as the Philippines and Indonesia would one expect to see an increase in overall trade openness as the number of education commitments increase. For countries at high- and low-income levels, results were insignificant (see Table 2).

### *What Characteristics Are Associated With Countries That Have Made Education Commitments?*

Descriptive statistics for the variables included in the probit regression are reported in Table 1 for all observations in the sample ( $N = 162$ ). According to results from the probit regression, the greatest single indicator of education commitments comes from the reduction of barriers to trade, such as unfavorable tax restrictions, needs tests, visa and work permit requirements, and citizenship

TABLE 2  
Descriptives for Variables Included in Probit Regression

<i>Variable Label</i>	<i>Variable Description</i>	<i>N</i>	<i>M</i>	<i>SD</i>
EDBAR	Presence of education barriers	162	7.4691	1.21927
WOPEN	General trade openness	163	85.9658	43.30192
FORENR	Foreign enrollment as % of overall enrollment	163	1.8613	5.40412
GECON1	Geography control: Land area in square miles	163	745378.8	2013437
GECON2	Geography control: Arable land % total	163	15.8013	13.81265
POPCON1	Population control: Population per square mile	163	169.0736	537.41301
POPCON2	Population control: Population as % of total	163	54.7829	22.81294
EDCON1	Education control: Literacy rate, adult total	163	79.68746	15.975207
EDCON2	Education control: Primary completion rate	163	75.4022	21.01235
INCON1	Income control: GNI per capita	163	5876.0000	8691.442
INCON2	Income control: GDP per capita as growth %	163	2.4984	3.68225

*Note.* GNI = gross national income; GDP = gross domestic product.

TABLE 3  
Probit Estimates of Variables Affecting Presence of Education Commitments

Variable	Model 1	Marginal Effect
Presence of education barriers	-.97043 (-5.66176)***	0.26065
General trade openness	.00416 (1.23103)	—
Foreign enrollment as % of overall enrollment	.01906 (.80982)	—
Geography: Land area in square miles	.00000 (1.64955)	—
Geography: Arable land % total	.01285 (1.14213)	—
Population: Population per square mile	-.00078 (-.75725)	—
Population: Population as % of total	-.00746 (-1.13972)	—
Education: Literacy rate, adult total	.00073 (.07547)	—
Education: Primary completion rate	-.00164 (-.24639)	—
Income: GNI per capita	.00001 (.52993)	—
Income: GDP per capita as growth %	-.05828 (-1.39599)	—
Constant	6.60693***	—

*Note.* Dependent variable is whether or not a country has made a commitment to Education Services under the General Agreement for Trade in Services. All values reported are for 2000. GNI = gross national income; GDP = gross domestic product.

\*\*\**p* < .001.

requirements. That is, a 1-unit decrease in the presence of these trade barriers is responsible for a .26 or 26% increase in the likelihood of a country having made commitments to education services trade, controlling for geographic, population, education, and income variables. The *t* statistic for this result is significant at *p* ≤ .001.

Coefficients for general trade openness and foreign enrollment produced positive impacts on the likelihood of education trade commitments; however, these coefficients were not significant at the .05 level. Other characteristics included in the model, such as general trade openness and foreign enrollment, do not have an identifiable impact on the likelihood of countries' having made education commitments. Full results of probit coefficients and their associated *t* statistics are reported in Table 3.

DISCUSSION

In answer to the research question What is the relationship between education services trade and overall trade?, results indicated that for lower-middle-income countries, education services trade commitments were positively correlated with higher levels of general trade openness. That is, as the number of education services trade commitments increases for a particular country, one finds a corresponding increase in that country's value of imports plus exports, divided by GDP. Although these results represent only the 2000–2001 calendar year (the most recent and



complete data available), it is assumed that this relationship would also be visible in subsequent years of data collection. As more recent observations become available through OCED, UNESCO, WTO, and World Bank data collection efforts, it will be useful to repeat this analysis with time-series data.

No significant association was found for high-income, upper-middle-income, and lower-income countries. This is not surprising, given that so few countries in the upper-middle and lower income brackets have made education services trade commitments. For high-income (OECD and non-OECD) countries, this lack of effect is also consistent with recent research on the effects of WTO membership on general trade openness, where findings indicated no significant relationship between membership in the WTO and overall trade openness (Rose, 2002).

The presence of a positive correlation between education services trade commitments and general trade openness is consistent with emergent education trade activities and policymaking in many of the 46 lower-middle-income countries. In countries like Brazil, the Dominican Republic, Turkey, China, Thailand, and Indonesia, significant efforts are underway to understand the potential opportunities in expanded education services trade and to construct mechanisms for deploying new modes of learning. For these countries, education trade would seem to mirror a larger trend toward increased marketization and privatization in all facets of the economy.

As an example, consider several lower-middle-income countries in southeast Asia. The Philippines, Indonesia, and Thailand, though largely lacking education commitments under GATS at this stage of the WTO negotiations, are each involved in initiatives designed to increase trade in education services.

Thailand has aggressively pursued liberalization in recent years, in part because of increasing access for its growing postsecondary student population. In 2003, only 27.4% of eligible Thai students were enrolled in higher education. Thailand has negotiated a variety of initiatives, including twinning arrangements such as an undergraduate double-degree program in tropical agriculture between Kasetsart University, Melbourne (Australia)-based Victoria University, and the American School of Bangkok, which provides an internationally focused undergraduate program licensed by the Thai Ministry of Education (Sadiman, 2004).

Recent policies have also resulted in favorable conditions for education services trade in Thailand. Under the Thailand-Australia Free Trade Agreement (2004), Thai secondary and higher education services could operate in Australia in all modes of supply except Mode 2. In turn, Australian higher education services operating in Thailand are limited to programs in life science, biotechnology, and nanotechnology and must be situated outside metropolitan areas. This arrangement represents an exciting type of bilateral agreement that uniquely positions Thailand to take advantage of market forces in Australia while expanding national access in areas of great need (Sadiman, 2004).

Funding cuts as a result of the 1988–1989 economic crisis forced Indonesia to examine outside sources for the provision and funding of education services. Today, more than 56% of funding for tertiary education comes from private sources (Fredrikkson, 2004). Indonesian students can attend the courses at the University of Phoenix over the Internet, and the Ministry of Trade is fully committed to opening markets to education services trade over time. Twinning agreements are occurring at both the intranational level (a partnership between the relatively new University Al-Azhar Indonesia and the Bandung Institute of Technology) and at international levels (as in the Netherlands Education Center in Indonesia's offering some 1,150 tertiary study programs; Sadiman, 2004).

In the Philippines, no specific tertiary education programs are underway as a result of GATS. However, a large number of professional and technical schools have been created in recent years to train health care and nursing professionals, both nationally and internationally. In addition, recent concern has arisen about the presence of diploma mills, programs of dubious quality that have consistently failed the Professional Regulation Commission quality assurance exam. Unfortunately, the presence of such programs is likely a temporary by-product of increased openness in educational services trade, until regulation and competition weed out most subpar providers. It is hoped that Thailand might serve as a model to the Philippines as the country expands its nongovernmental education offerings from technical and professional training increasingly toward alternate methods of traditional tertiary education.

What does this data mean? At a minimum, that southeast Asian lower-middle-income countries, and others like them, are working aggressively to open their borders to education trade and that, although cultural and social concerns about the unique nature of education may have some relevance, they are not the criteria upon which countries are making decisions. Although innovating in response to national demand, these countries have also recognized a market for their services outside local borders. Education could quite possibly be a unique provision, but practically speaking for these lower-middle-income countries, education efforts are following the marketization trends seen in nearly every sector of an increasingly global economy.

Is education unique? Is education trade subject to different parameters than overall trade? What is the relationship between education and the market? This study indicates that for at least the lower-middle-income countries, education trade is not different from general trade. However, further research into all levels of education and studies using powerful, multivariate methods will provide the most comprehensive picture of trade behavior.

When considering the question, How is education services trade related to overall trade?, results indicated that when controlling for demographic factors such as national population, geography, income, and educational attainment, education barriers produced a moderate effect on overall trade openness (i.e., countries with

TABLE 4  
Sample Descriptive Statistics for Educational  
Openness Index

<i>N</i>	<i>Mid</i>	<i>Max</i>	<i>M</i>	<i>SD</i>
22	.30	.68	.51	.081

fewer education barriers had, on average, slightly higher levels of general trade openness).

In further investigating this relationship through a probit regression technique, it was determined that the presence of education barriers was responsible for a 26% marginal effect on the likelihood that a country holds at least one education services commitment under GATS. Neither foreign enrollment and overall trade openness nor the control variables included in the model had significant marginal effects on the regression or probit outcomes.

Again, this finding is consistent with existing literature dealing with GATS and its potential impact on education services trade. Limiting the presence of barriers to trade was identified early in the negotiations as an essential goal of progressive liberalization (WTO Secretariat, 1998), and more recently, studies have attempted to measure and quantify the impact of these barriers (Kemp, 2001; Nguyen-Wells & Hong, 2000). In each case, researchers have pointed to difficulties in data collection and the role of future researchers in extending the models and methodologies represented in their work to estimate the impact of these barriers on countries' educational markets—including cost, quality, and public expenditure—as well as on longer term measures of economic health.

A final area of interest is the creation of an Educational Openness Index. In an effort to add to recent attempts to focus attention on quantifying education trade (Kemp, 2001; Larsen et al., 2002; Nguyen-Hong & Wells, 2000), three variables were transformed into an index designed to judge specifically countries' relative "openness" as related to the cross-border movement of education services. Sample results from this index are presented in Tables 4 and 5.

As mentioned previously, the limited sample size ( $n = 24$ ) made statistically useful results from Educational Openness Index analyses difficult to provide. In addition, no comparisons across income levels were possible because of sample size. Although such an Index would doubtlessly provide more robust information about the strength of education services trade, results indicated that because of lack of available data measures, such calculations are premature.

Given the limited data available and lack of consistent measures of reporting across countries and regions, it is imperative that results of this and all analyses in this study be considered preliminary and of limited generalizability, particularly for non-OECD-developed and developing countries, for which data are particularly

TABLE 5  
Sample Education Openness Indexes for Selected Countries

Country	EDBAR	EDCOMW	FORENR	EOINDEX
Australia	3.00	.69	5.64	.43
Austria	3.00	.45	8.75	.51
Belgium	3.00	.85	34.25	.68
Czech Republic	2.50	1.00	2.37	.52
France	3.50	.85	21.30	.56
Japan	6.50	.85	2.00	.30
Mexico	2.50	.85	.60	.45
Slovak Republic	2.50	1.00	1.40	.51

difficult to obtain. More research is needed to determine the exact nature of these results as well as their impact over time.

Limitations and Areas for Further Analysis

As with any study of education trade, missing and incomplete data are the primary limitation of this analysis. For example, the EOINDEX variable could be calculated for only 22 of the 44 countries that have committed to opening their education services markets under GATS. In addition, missing data for countries across data cycles required that in several cases data substitution measures were necessary. The full version of this article includes detailed technical notes describing these substitutions.

Rigorous analysis of education trade is limited by the data collected at national and international levels, particularly regarding collection itself. A large percentage of least developed and developing countries do not collect and report even basic cross-national education statistics. Data collection is time-intensive and expensive,

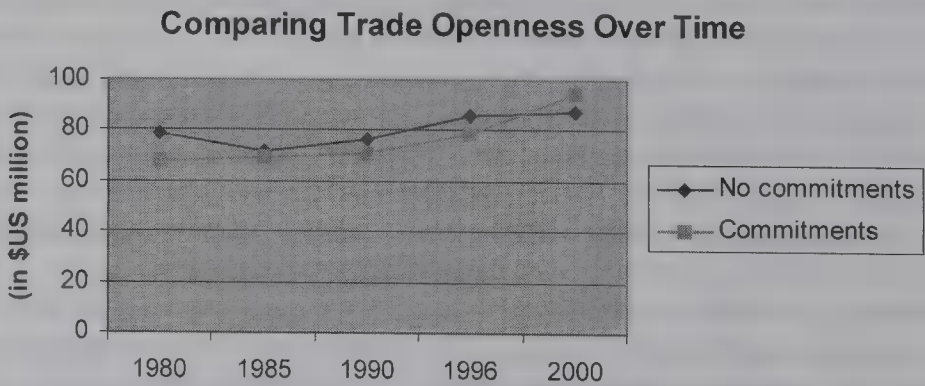


FIGURE 7 Comparison of trade openness between countries with/without commitments under GATS.



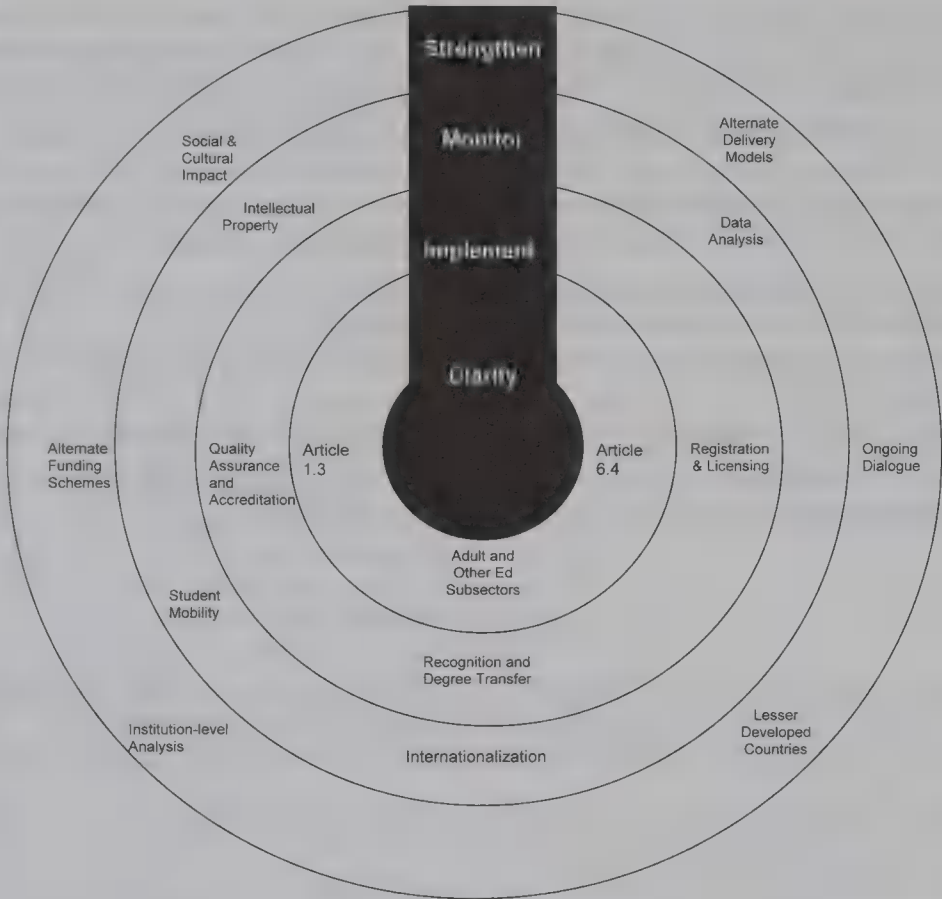


FIGURE 8 A systemic model for improvement in education services trade (Payne, 2005).

and particularly in the case of developing countries, time and attention often (and rightly) fall to domestic education issues over cross-border trade. Better measures are needed to ensure accuracy in data analysis, and more structured collection methods can strengthen the quality of currently available data (Knight, 2002b, 2003; Larsen et al., 2002; Nguyen-Hong & Wells, 2000, and others).

Second, most countries do not segment their import and export of education goods, programs, and services from their overall trade statistics. In their WTO–GATS proposals, Australia, the United States, Japan, and New Zealand call for involvement from other nations in better tracking of education trade measures. Statistics such as numbers of foreign students by country of origin, education goods and services as percentages of overall import and export, percentage of private versus public spending on education, and amount of spending on lifelong learning and education programs do not exist in aggregate today, even for many developed countries. It is recommended that these variables be adjusted to reflect the four modes of education services supply under GATS (Kemp, 2001; Knight, 2002a, 2003; Larsen et al., 2002; Nguyen-Hong & Wells, 2000; Sauve, 2002).

Regarding directions for future research, the model described in Figure 7 represents the synthesis of countless recommendations related to better understanding of the impact of GATS on the education-attaining public. Designed to flow from items of critical short-term importance outward to longer term, ongoing areas of inquiry, this model segments key areas of policy opinion and analysis into four major recommendations: clarification, implementation, modification, and strengthening (Payne, 2005). It is my intention that this model serve as a foundation for ongoing research and study related to the increasing focus on cross-border movement of educational resources, goods, services, and materials.

Is education unique, or is it subject to the same market forces as transportation, textiles, and other trade sectors? It is likely too soon to tell. However, for those countries that are considering making commitments to reduce barriers to trade, these early findings may provide one avenue for analyzing the relative threats and opportunities of liberalizing access to education programs.

## REFERENCES

- American Council on Education. (2004). An overview of higher education and the GATS. Retrieved November 5, 2004, from <http://www.acenet.edu/programs/international/gats/overview.cfm>
- Altbach, P. (2001, May 11). Why higher education is not a global commodity. *Chronicle of Higher Education*, 47, B20.
- Altbach, P. (2002, Summer). Knowledge and education as international commodities: The collapse of the common good. *International Higher Education*.
- Altbach, P. (2003, Spring). The United States and international education trade. *International Higher Education*, Available at [www.bc.edu](http://www.bc.edu).
- Ascher, B. (2001, May). *Education and training services in international trade agreements*. Paper presented at US-OECD Trade Forum, Washington, DC.
- Asia-Pacific Economic Cooperation. (2001). *Measures affecting trade and investment in education services in the Asia-Pacific region* (Report to the APEC group on Services 2000). Singapore: Author.
- Birdsall, N., & Hamoudi, A. (2002). *Commodity dependence, trade, and growth: when "openness" is not enough* (Center for Global Development Working Paper No. 7). Center for Global Development.
- Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall.
- Chen, P., & Popovich, P. (2002). *Correlation: Parametric and nonparametric measures*. Newbury Park, CA: Sage.
- Center for Quality Assurance in International Education. (1999). *Foundations of globalization of higher education in the professions*. Washington, DC: Author.
- Dollar, D. (1992). Outward-oriented developing economies really do grow more rapidly: Evidence from 95 LDCs, 1976–85. *Economic Development and Cultural Change*, 40, 523–544.
- Dollar, D., & Kraay, A. (2001). *Trade, growth, and poverty* (World Bank Working Paper No. 2615). Washington, DC: World Bank.
- Edwards, S. (1997). *Openness, productivity, and growth: What do we really know?* (NBER Working Paper Number 5978). National Bureau of Economic Research.
- Fredriksson, U. (2004). *Quality education: The key role of teachers*. Brussels, Belgium: Education International.
- Greenaway, D., Morgan, W., & Wright, P. (2002). Trade liberalization and growth in developing countries. *Journal of Development Economics*, 67, 229–244.

- Heston, A., Summers, R., & Aten, B. (2002). *Penn World Table, version 6.1*. Philadelphia, PA: Center for International Comparisons at the University of Pennsylvania.
- Heyneman, S. (2001). The growing international commercial market for educational goods and services. *International Journal of Educational Development*, 21, 345–359.
- Heyneman, S. (2003). International education: A retrospective. *Peabody Journal of Education*, 78(1), 33–53.
- Hill, D. (2001). Global capital, neo-liberalism, and privatization: The growth of educational inequality. In D. Hill & M. Cole (Eds.), *Schooling and policy: Fact, concept, and policy* (pp. 25–54). London: Kogan Page.
- Hoekman, B. (1995). *The WTO, the EU and the Arab world: Trade policy priorities and pitfalls* (Center for Economic Policy Research Discussion Paper 1226). London: CEPR.
- Jarvis, P. (2000). Globalization, the learning society, and comparative education. *Comparative Education*, 36, 343–355.
- Kemp, S. (2001). Trade in education services and the impacts of barriers to trade. In C. Findlay & T. Warren (Eds.) *Impediments to trade in services: Measurement and policy implications* (pp. 231–244). New York: Routledge.
- Knight, J. (2002a). Trade creep: Implications of the GATS for educational trade policy. *International Higher Education*, 28(2), available at [www/bc/edu](http://www/bc/edu)
- Knight, J. (2002b). *Trade in higher education services: The implications of GATS*. London: The Observatory on Borderless Higher Education.
- Knight, J. (2003, Spring). Trade talk—the four modes. *International Higher Education*.
- Konen, D., & Maskus, K. (2004). *Quantifying the impact of services liberalization in a developing country* (World Bank Policy Research Working Paper 3193). Washington, DC: Author.
- Larsen, K., Morris, R., & Martin, J. (2002). Trade in educational services: Trends and emerging issues. *World Economy*, 25(6) 3–18.
- Lenn, M. P. (1999). *Higher education and the global marketplace: A practical guide to sustaining quality*. Washington, DC: Center for Quality Assurance in International Education.
- Lenn, M. P. (2000). *The globalization of the professions in the United States and Canada: a survey and analysis*. Washington: The Center for Quality Assurance in International Education.
- Lenn, M. P., & Miller, B. (2000). *The foundations of globalization of higher education and the professions*. Washington, DC: NCITE.
- McGuire, B., & Schuele, M. (2000). Restriction of international trade in maritime services. In C. Findlay & T. Warren (Eds.), *Impediments to trade in services: Measurement and policy implications* (pp. 201–214). London: Routledge.
- McGuire, B., Schuele, M., & T. Smith. (2000). Restrictiveness of international trade in banking services. In C. Findlay & T. Warren (Eds.), *Impediments to trade in services: Measurement and policy implications*. (pp. 172–188) London: Routledge.
- Naidoo, R. (2007). *Higher education as a global commodity: The perils and promises for developing countries*. London: Observatory on Borderless Higher Education.
- Nguyen-Hong, D., & Wells, R. (2000). *Restrictions on trade in Education Services: Some basic indices*. Canberra, Australia: Productivity Commission Staff Research Paper.
- Nyborg, P. (2002). *GATS in the light of increasing internationalisation of higher education: Quality assurance and recognition*. Paris: OECD.
- Organisation for Economic Cooperation and Development. (2002). *Current commitments under the GATS in Education Services* (Paper prepared for the OECD/US Forum on Trade in Education Services). Washington, DC: OECD/CERI.
- Pampel, F. (2000). *Logistic regression: A primer*. Thousand Oaks, CA: Sage.
- Payne, B. (2005). *Influences and implications of international trade in education services*. Unpublished doctoral dissertation, Vanderbilt University.
- Pritchett, L. (1994). Measuring outward orientation in LDCs: Can it be done? *Journal of Development Economics*, 49, 307–335.

- Rose, A. (2004, March). Do we really know that the WTO increases trade? *American Economic Review*, 94(1), 98–114.
- Sachs, J., & Warner, A. (1992). Economic reform and the process of global integration. *Brookings Papers on Economic Activity*, 1, 1–117.
- Sachs, J., & Warner, A. (1995). Fundamental sources of long-run growth. *American Economic Review, Papers and Proceedings*, 87, 184–188.
- Sadiman, A. (2004, August 23). *Trade in Educational Services in southeast Asia: Opportunities and challenges*. Plenary paper presented at the Beijing Forum on Trade and Cooperation in Education Services in Asia.
- Sauve, P. (2002). *Trade in education and the GATS: What's in, what's out, what's all the fuss about?* Background paper prepared for the OECD/US Forum on Trade in Education Services, Washington, DC.
- Sidhu, R. (2007). GATS and the new developmentalism: Governing transitional education. *Comparative Education Review*, 51, 203–229.
- Thailand–Australia Free Trade Agreement. (2004). New South Wales: Australian Department of Foreign Government and Trade.
- U.S. Department of Commerce. (1998). *U.S. international transactions in private services, a guide to the surveys conducted by the Bureau of Economic Analysis*. Washington, DC: Author.
- U.S. Department of Commerce. (2000). *U.S. international services—Cross-border trade in 1999 and sales through affiliates in 1998, survey of current business*. Washington, DC: Author.
- U.S. International Trade Commission. (2006). *Recent trends in U.S. services trade: 2006 annual report*. Washington, DC: Author.
- Van Den Wende, M. (2001). Internationalism policies: About new trends and contrasting paradigms. *Higher Education Policy*, 14, 249–259.
- World Bank. (2004). *World development report 2005: A better investment climate for everyone*. Washington, DC: Author.
- World Trade Organization. (1999a). *The General Agreement on Trade in Services (GATS): Objectives, coverage and disciplines* (Document prepared by the WTO Secretariat).
- World Trade Organization. (1999b). *Market access: Unfinished business—post Uruguay Round inventory and issues* (Special Study, No. 6). Geneva, Switzerland: Author.
- WTO Secretariat. (1998a, September 23). *Education Services* (S/C/W/49).
- WTO (2001, April). Trade in Services Division. The general agenda for trade in services: Objectives, coverage, and disciplines. WTO.
- WTO Secretariat (2001). GATS: Fact or fiction. Retrieved December 13, 2007 from [http://www.wto.org/english/gatsfacts1004\\_e.pdf](http://www.wto.org/english/gatsfacts1004_e.pdf)

## APPENDIX

### TECHNICAL NOTES REGARDING DATA COLLECTION AND TRANSFORMATION

#### A. Procedure for developing master WOPEN dataset:

1. Generate overall list of countries from PWT 6.1
2. Assign values for education commitments based on WTO Online Database.
3. Exclude countries for which no openness or education trade commitment exists (45):



Aruba, Andorra, Afghanistan, Netherlands Antilles, American Samoa, Bahamas, Bosnia and Herzegovina, Bermuda, Bhutan, Cambodia, Channel Islands, Cayman Islands, Eritrea, Faeroe Islands, Micronesia, Greenland, Guam, Isle of Man, Iraq, Kiribati, Laos, Liberia, Libya, Monaco, Marshall Islands, Northern Mariana Islands, Mayotte, New Caledonia, Oman, Palau, Puerto Rico, French Polynesia, Saudi Arabia, Samoa, Sudan, San Marino, Somalia, Turkmenistan, Tonga, Taiwan, Uzbekistan, US Virgin Islands, Vietnam, Vanatu, West Bank and Gaza.

4. In cases where WOPEN exists but no information can be found on education commitments, assign value of 0 (no commitment) to country (26):

Albania, Algeria, Armenia, Azerbaijan, Belarus, Cape Verdi, Comoros, Croatia, Ethiopia, Georgia, Iran, Jordan, Kazakhstan, Lebanon, Lithuania, Moldova, Macedonia, Nepal, Russia, San Tome and Principe, Seychelles, Syria, Tajikistan, Ukraine, Yemen, Yugoslavia.

5. In cases where WOPEN does not exist but education commitments do, correct for missing data by substituting the mean WOPEN measure for the World Bank economic indicator associated with the missing country (28):

Low income ( $M = 68.87023231$ )

Angola, Central African Republic, Haiti, Myanmar, Mongolia, Mauritania, Papua New Guinea, Democratic Republic of Korea, Sierra Leone, Solomon Islands, Congo DP

Lower-middle income ( $M = 67.07508853$ )

Djibouti, Fiji, Guyana, Namibia, Suriname, Yugoslavia

Upper-middle income ( $M = 109.6473437$ )

Botswana, Maldives

High income ( $M = 119.2172347$ )

United Arab Emirates, Bahrain, Brunei, Cyprus, Kuwait, Liechtenstein, Malta, Qatar, Singapore

# Diagnosis, Treatment, and Educational Implications for Students With Attention-Deficit/Hyperactivity Disorder in the United States, Australia, and the United Kingdom

Sarah Schlachter  
*Vanderbilt University*

For many years there has been debate over Attention-Deficit/Hyperactivity Disorder (ADHD) and whether this condition, which commonly afflicts adolescent children, is a medical or social condition and whether it is exclusively an American phenomenon. This article reviews the basis of ADHD's definition, diagnosis, treatment, and educational implications across three countries: the United States, Australia, and the United Kingdom. The differences in approach have clear and significant consequences for children and their futures.

It is fairly likely that if you asked the average person on the street in the United States if they have heard of Ritalin or of an illness called Attention-Deficit/Hyperactivity Disorder, also known as ADD or ADHD, that person would say yes. Opinions on ADHD range from it being a made-up disorder used as an excuse for low-achieving students to it being a debilitating illness with the potential to severely limit the academic prospects of young students. For nearly 30 years, this illness and its medical treatment have been prominent in discussion of the state of children and of education today.

An examination of literature from the United States, Australia, and the United Kingdom demonstrates that this same range of opinions can be found among scholars in education and medical journals. In an effort to break down the veracity of these opinions, this article investigates the definitions and criteria used to

ascertain a diagnosis of ADHD. It then compares and contrasts the rates of prevalence that characterize the populations of the three countries. Finally, it considers the methods of treatment utilized to negate ADHD's effects and symptoms, including the use of medication and behavioral strategies for the classroom and beyond.

## DEFINITION AND CRITERIA

The standard definition and criteria for the diagnosis of ADHD in the United States comes from the American Psychiatric Association's (1994) *Diagnostic and Statistical Manual of Mental Disorders (DSM)*. This manual has undergone multiple updates since its first publication in 1952. The most recent edition available is the *DSM-IV-TR*, which was published in 1994. The evolution of the labels, criteria, and symptoms that have surrounded ADHD can easily be inferred through the six full pages devoted to the disorder in the *DSM-IV-TR*. In fact, it has been said that "no other childhood psychopathology has undergone as much renaming and reconceptualization as the hyperactive disorder" (Gomez, Harvey, Quick, Scharer, & Harris, 1999, p. 265).

The *DSM-IV-TR* has expanded the symptoms and criteria of ADHD, and the definition now includes three subtypes: ADHD Predominantly Hyperactive-Impulsive type, ADHD Predominantly Inattentive Type, and ADHD Combined Type. To be diagnosed as either ADHD Predominantly Inattentive Type or ADHD Predominantly Hyperactive-Impulsive, a person must exhibit symptoms for at least 6 months "to a degree that is maladaptive and inconsistent with developmental level." Each subtype has a list of 9 symptoms, and 6 of those must be present for a diagnosis. ADHD Combined Type requires 6 symptoms out of the possible 18 for the same length of time and extent (See Appendix A).

Another definition and set of criteria that are more commonly used across Europe come from the World Health Organization's (WHO's) International Classification of Diseases (ICD). The 10th version of the ICD (ICD-10) has diagnostic criteria for ADHD that are remarkably similar to those of the *DSM-IV-TR*, including the listings of possible symptoms, with a diagnosis requiring at least 6 of 10 such symptoms present in the child for at least 6 months, also "to a degree that is maladaptive and inconsistent with the development of the child." The ICD-10 also includes the same three subtypes of inattention, hyperactivity, and impulsivity (See Appendix B).

As with the *DSM-IV-TR*, there are frequent updates and clarifications between editions. The term used in ICD-10 for ADHD is *hyperkinetic disorder*. Although hyperkinetic disorder is not exactly the same as ADHD, the term ADHD is still commonly used in British studies, perhaps for its universality, and this article uses the term ADHD. The ICD-10 was approved in 1990 and went into use by WHO member states in 1992.

Although nearly identical, there are some key differences in the definitions, which ultimately lead to very different patterns of diagnosis. For example, the ICD-10 requires that the pervasiveness and persistence of symptoms be present in at least two situations (such as home and school). The *DSM-IV-TR* would allow a diagnosis of ADHD with the presence of symptoms in only one situation. More significantly, within the hyperactivity subtype the *DSM-IV-TR* criteria allow a diagnosis if the child displays *either* impulsiveness *or* inattention, whereas the ICD-10 requires both symptoms (Reason, 1999). As is discussed in the comparison of prevalence rates among countries such as the United States and Australia (which primarily use the *DSM-IV-TR* as a basis) versus Great Britain (which primarily uses the ICD-10), the slight variances in criteria play a large role in the predominance of the disorder in populations.

## DIAGNOSIS

Despite the *DSM-IV-TR* and ICD-10 standards, consistent diagnosis of ADHD remains difficult for a variety of reasons. First, there is little regularity on who is making the diagnoses of the disorder. A variety of medical professionals such as general practitioners, pediatricians, or mental health specialists may be assessing the subject. Prior to a visit with a doctor, the student will be in regular contact with a number of other individuals who may play a large role identifying the disorder and bringing about a diagnosis. These individuals could include parents, teachers, coaches, and other caregivers. Despite the large role these people play in a child's life, there is still relatively little room for their input into a formal diagnosis. "To date, there are no descriptive data for parent and teacher ratings of AD/HD symptoms listed in *DMV-IV*" (Gomez et al., 1999, p. 267). Although a medical diagnosis can only officially be made by a doctor, other actors can and should play a major role in defining the child's illness.

In some ways, the *DSM-IV-TR* criteria lack the specificity necessary to function as working guide for diagnosis. "The current *DSM-IV* edition can equally be criticized for not providing clear indications of abnormal levels for the symptoms listed" (Gomez et al., 1999, p. 267). This lack of specificity introduces a high level of subjectivity into the diagnosis process. Such high level of subjectivity will necessarily affect rates of prevalence and cloud an accurate picture of the illness' scope.

## PREVALENCE

### *United States*

A direct result of the difficulty in diagnosing ADHD is a wide variation in prevalence of the disorder in children. The *DSM-IV-TR* puts the prevalence rate at



between 3% and 5%. Within this group, the vast majority of those diagnosed are male, with some studies suggesting that up to 90% of cases of attention deficit are boys (Purdie, Hattie, & Carroll, 2002). Although there is increasing acceptance that ADHD can follow children into adulthood, the disorder is still generally accepted to be one that afflicts young children up to adolescence, roughly between the ages of 7 and 12 (Ciechomski, Blashki, & Tonge, 2004).

A tremendous increase in the number of publications and studies examining ADHD in the United States has resulted in statistics that reflect great differences in prevalence rates cited. Overall, other stated rates exceed the prevalence cited as the standard by the *DSM-IV-TR*. The American Academy of Pediatrics puts the rate between 4% and 12% (American Academy of Pediatrics, 2001). Other studies have put the figure between 20% and 24% (Purdie et al., 2002).

Many issues contribute to the difficulty in determining a generally accepted rate of prevalence for the disorder. The most frequently cited explanation for the variation is the existence of different definitions and criteria for diagnosing ADHD. As a result, the rates of occurrence will likely be skewed. Although the *DSM-IV-TR* is a generally accepted standard, there remain areas within that standard that are unclear. Moreover, as has already been discussed, the issues of multiagency in diagnosing means that all parties may not use the *DSM-IV-TR*, let alone interpret it and apply it universally. Third, variations in study methodology will play a major role in affecting rates and figures. Differences in populations examined will skew the numbers significantly. Finally, it is generally accepted that ADHD has a very high rate of comorbidity with other disorders or illnesses. According to one study, 65% of ADHD diagnosed children have another diagnosed psychiatric or behavioral issue (Shaw, Wagner, Eastwood, & Mitchell, 2002). The concurrence of illnesses could result in masking symptoms or misinterpreting symptoms for one disorder or another.

### *Australia*

As in the United States, there has been a belief in Australia that the occurrence of ADHD has grown very quickly and perhaps with little substantiation. Factors and details surrounding the diagnosis of ADHD in Australia are extremely similar to those in the United States.

Most significantly, the standard definition and criteria for diagnosis in Australia of the disorder is also the American Psychiatric Association's *DSM* and its subsequent revisions. Literature from Australia indicate the same basic problems and limitations resulting from this definition. There were multiple indications that the role of data from parents and teachers was lacking, "a multidimensional approach whereby information is gathered from a number of sources (e.g. Parents, teachers) is regarded as best practice" (Ciechomski et al., 2004, p. 1000). A 1997 report from the National Health and Medical Research Council of Australia

(NHMRC) recommended that parent and teacher input play a greater role in diagnosis, though this may have a serious effect on prevalence rates. As was stated in the NHMRC report, "Even small differences in diagnostic procedures can affect rates, which in turn have a powerful effect on the predictive value of diagnostic tests." (p. 22)

Because it is difficult to obtain reliable rates of prevalence, it is also difficult to soundly compare nations in the frequency of ADHD. One study put the worldwide prevalence of ADHD at between 1.7% and 6.7% (Shaw et al., 2002). Various studies of Australian children have found the prevalence rate to be within or very near the quoted 3% to 5% for U.S. children found in the *DSM-IV-TR*. The rates quoted in the 1997 NHMRC report were between 2.3% and 6% for the child population of Australia as a whole. Because it is a given that the figures will vary widely, some different perspectives on the Australian rates and more focused views could provide greater insight.

One unique way to look at prevalence rates is to consider how often patients are seen and/or diagnosed in doctors' offices. Although these data are also subject to some inconsistency in diagnosis because of definition, it demonstrates another perspective and it appears to indicate that ADHD may be underdiagnosed. In a study that looked at rates of usage of medical and school-based services, only 28% of students with symptoms of ADHD sought help, with 41% going to medical services, 39% going to school services, and 20% to both (Sawyer et al., 2004). The rate at which Australian general practitioners see children with ADHD was seemingly low, with only between one and five cases per year out of an average of more than 250 children seen per year (Chiechomski et al., 2004).

### *United Kingdom*

As with the United States and Australia, the prevalence rates of children diagnosed with ADHD (as used in the ICD-10) vary, and such rates are dependent on the highly subjective nature of diagnosis. One article claims that between just 0.5% and 1% of children age 7 and younger in Great Britain have ADHD. The generally accepted rate of occurrence is around 1% to 2% of children (Parr, Ward, & Inman, 2003). This low level of occurrence is likely the result of multiple factors. First, the British usage of the more exclusive WHO definition of hyperkinetic disorder means that fewer children will fulfill the symptomatic requirements of the diagnosis. It is likely that were the British to employ the *DSM-IV-TR*'s wider standards for ADHD, British rates of prevalence for ADHD would be significantly higher than their current rates for hyperkinetic disorder.

Second, the British view of the impairment is extremely different from that in the United States and Australia, and this affects the likelihood of children to be diagnosed. As can be inferred by the usage of the ICD-10 definition, ADHD is not a term that is liberally applied to children. "In Britain, ADHD is conceptualized

as a psychosocial problem whereas in America ADHD is viewed as a medical problem" (Reid & Maag, 1997). This is coupled with the British view that each child who may be afflicted by ADHD is unique and that the particular symptoms and presentation of the illness will be different in every case.

It seems that there is an overall reluctance to apply the label of ADHD to a child in Britain because this may negatively affect the child's school or social life as a result. This contrasts starkly to the United States, where parents may actively seek a diagnosis of ADHD to potentially secure extra assistance or at least understanding for the child in social and educational settings. Indeed, one article suggests that ADHD is another illness popular in the United States because of "America's propensity toward glorifying victimization" (Reid & Maag, 1997).

Finally, the overall lesser acceptance of ADHD and more infrequent occurrence may actually be contributing to a reality of underdiagnosis in the United Kingdom. Although many studies have speculated that ADHD is generally underdiagnosed all over the world, the United Kingdom may be particularly at risk because of the structure of its health system. National treatment guidelines state that for a diagnosis of ADHD, a child must see a specialist, which would be either a pediatrician or a child psychiatrist. The majority of parents and children typically come into contact only with their general practitioners, who are both unauthorized to diagnose a hyperactivity issue and likely ill-prepared to recognize the symptoms. As British parents are unexposed to and less familiar with ADHD, one study showed that parents who may have concerns about their child's behavior seek advice only from education professionals and frequently do so stating that the problem is potentially a learning disorder rather than a mental illness (Sayal, Goodman, & Ford, 2006). Education professionals in the United Kingdom are similar to the general practitioners in that they are likely ill-prepared and undereducated about ADHD and may not direct parents to the appropriate resources.

Although fears abound in the United States of overdiagnosis of ADHD, and those fears are beginning to spread to Australia, it appears that the United Kingdom is understating the case among its children. Indeed, a comparison of the phenomenon of ADHD worldwide states, "There is no convincing difference between the prevalence of this disorder in the USA and most other countries or cultures." Moreover, "the apparent 20-fold difference in the prevalence of hyperactivity reflects differences in the definition of the condition rather than real differences in behavior" (Faroane, Sergeant, Gillberg, & Biederman, 2003, p. 104).

In all three contexts, there are many factors at work that complicate the situation, not the least of which are the social factors. The social constructions of and assumptions about ADHD have grown in the past 2 decades alongside the numbers of children diagnosed. These assumptions and preconceptions can play a large role in the diagnosis of ADHD when those without medical training, such as teachers and parents, allow their preconceptions to affect their involvement in the

diagnosing of ADHD. "Notions of what constitutes normal classroom behavior have led to the application of the label ADHD" (Purdie et al., 2002, p. 65).

## TREATMENT

There are clearly major differences in how different countries approach the diagnosis of ADHD, so it is not surprising that there are major differences in how it is treated, both medically and behaviorally. Stimulant medications such as methylphenidate, better known by its brand name, Ritalin, and dextroamphetamine are often prescribed as a means of increasing children's ability to focus. Behavioral modification strategies, especially those employed in the classroom, are often recommended in accompaniment to medication, though these appear to be less frequently employed than medication alone. Overall, there is some consensus that treatment should be multimodal, but studies to show the efficacy of this approach are limited and actual treatment practices do not necessarily currently reflect multimodal recommendations.

### *United States*

In the United States today, there is a general impression that an excessive number of children are diagnosed with ADHD and that they are subsequently overmedicated with stimulants that may or may not be necessary to improve their behavior. We have already discussed the veracity of the claim that American children are overdiagnosed with the disorder, and it seems that evidence supports the notion that they may also be overmedicated. A recent meta-analysis of ADHD diagnoses and treatment stated, "Medication is the most commonly reported form of intervention for children with ADHD" (Purdie et al., 2002, p. 66). Although medication is common, the limitations of its effects are also recognized. Medication will not "cure" a child and symptoms will persist, though perhaps to a lesser degree. Complete "normalization" will not be achieved. Medication, also, usually only has short-term effects.

Usage of psychotropic stimulants increased in the United States between 1987 and 1996 from 0.6% to 2.4%. Between 1997 and 2002, the increase was less severe, from 2.7% to 2.9%, or 2.2 million children (Zuvekas, Vitiello, & Norquist, 2006). Although the difficulties in comparing rates of stimulant usage are comparable to the difficulties in comparing prevalence rates, much evidence indicates that stimulant medication prescription in the United States varies greatly from other countries; "methylphenidate is prescribed at a considerably higher rate in the United States than in other developed nations" (Wolraich, 2003, p. 160).



In 2001, the American Academy of Pediatrics published its “Clinical Practice Guideline: Treatment of the School-Aged Child with Attention-Deficit/Hyperactivity Disorder” in an attempt to provide consistency of treatment. The number two recommendation in that guideline was “The treating clinician, parents, and the child, in collaboration with school personnel, should specify appropriate target outcomes to guide management” (p. 1033). This indicates the significant role schools should play in treating children with ADHD. It seems logical that treatment of ADHD include a strategy for the classroom because children with ADHD often have increased difficulty in school (Kos, Richdale, & Hay, 2006). According to Kos et al. there is a “dearth” of literature both of information for teachers currently in service and a lack of preservice training as well.

Typical behavioral strategies employed by teachers can be categorized as proactive and reactive. Proactive measures include choice-making interventions, peer tutoring, and computer-assisted instruction. Reactive measures are more common and have a greater history of usage in the classroom. These measures include verbal reprimand for distractive behavior, token reinforcement, and self-management interventions (DuPaul & Weyandt, 2006).

Many studies address the need for increased structure in the classroom both in terms of activities and the physical space of the classroom. Multiple sources indicate the desirability of a formal arrangement of desks and space. It is also supposedly more beneficial for students with ADHD to be seated near the front of the class and near the teacher as a means of keeping them on task. Noise levels should be reduced and frequent breaks should be incorporated into the structure of the day. In attempting to attend to students with ADHD in the classroom, teachers need to address all three aspects of ADHD—inattention, impulsivity, and hyperactivity—through the aforementioned techniques to achieve positive results (Purdie et al., 2002).

### *Australia*

As was the case with the definition of the ADHD and the general prevalence rates, the treatment and interventions generally employed for Australian children are very similar to those for American children. There is significant primary reliance on medication with comparable behavioral and classroom interventions as secondary strategies. Rates of medication use are similar to rates in the United States, though the difficulty in establishing reliable bases within studies for comparative purposes is also difficult.

Various studies showed that between 1.8% and 2% of school-age children in Australia used stimulant medication to address symptoms of ADHD between 2000 and 2002. The overall use of stimulants increased by 26% between 1984 and 2000, with an eightfold increase between 1994 and 2000. Relative to other countries,

the Australian rate of stimulant use is “only exceeded by the USA and Canada” (Isaacs, 2006, p. 545).

Also similar to the United States is the theoretical emphasis on multimodal treatment that does not seem to be reflected in the number describing treatment. The National Health and Medical Research (1997) stated, “A multi-modal approach, especially with educational and behavioural supports should be used if available” (p. 41). The high rates of medication usage suggest that multimodal treatment may not, in fact, be employed as frequently as the report suggests it should. “Behavioural intervention was underutilized despite its documented positive role” (Concannon & Tang, 2005, p. 625).

The advice to educators for classroom strategies meant to serve as behavioral interventions is extremely similar to that given to American educators. The NHMRC report discusses areas that should be addressed: maximizing attention and concentration, assisting the child in following instructions, reducing overactivity, countering impulsivity and inflexibility, improving socialization, and more. Each of these areas has specific actions such as physical classroom arrangement, allowing choice, maintaining a fixed routine, and allowing frequent breaks.

In comparison to the claims that U.S. educators have few formal resources and little training in teaching ADHD students, the South Australia Department of Education, Training and Employment has issued classroom behavioral strategies specific to students with ADHD. The strategies include positive reinforcement, negative consequences, emotional support, planned ignoring, and classroom organization. The environmental recommendations included making the classroom “active” and “quiet” (Kos et al., 2006).

### *United Kingdom*

In the United Kingdom, attitudes toward treatment in comparison to the United States and Australia are as dissimilar as attitudes toward diagnosis. Usage of stimulant medications is practiced in the United Kingdom but to a much lesser degree, and other options, such as behavioral interventions, are pursued more vigorously. Modification of classroom practices by teachers appears to be largely the same, though throughout the literature there was more discussion of the degree of the school’s role in treatment, as opposed to specific actions that could be taken.

Unlike in the United States and Australia, it was very difficult to find U.K. statistics on the usage of stimulants to treat ADHD. It seems that this may be because the United Kingdom has only recently begun to diagnose more cases of ADHD and there is therefore little history of treatment. Multiple studies indicated that prescriptions of stimulant medication for treatment of ADHD are increasing, consistent with an increase in diagnoses. “Despite its relatively late start compared to North American practice, paediatric psychopharmacology in the UK is now developing apace in terms of both clinical practice and evaluative research”

(Bramble, 2003, p. 176). Indeed, it is possible that the historical attitude toward ADHD and its merit as a disorder have affected the availability of research, as it is easy to see from a simple search of Medline and PsycINFO databases that much more data are available from the United States.

In just one article that was located, there was mention of the rate of stimulant medication usage in U.K. boys in 1999. This article stated that .53% of those studied were being treated with drugs, and it noted that treatment in the United Kingdom using stimulant drugs has been on the rise since the mid-1990s. This is opposed to the United States, where stimulants have been in use since the 1960s, and a study of a similar population to that of the U.K. study showed a 9.3% rate of drug treatment in the United States in 1995, 4 years prior to the U.K. study (Jick, Kaye, & Black, 2004).

Clearly, the acceptance of treatment by medication is less than that of the United States and Australia, and this is further demonstrated through discussions of other means of treatment in the United Kingdom. Although all three countries promote multimodal treatment of ADHD, the United Kingdom seems to be the only one to consistently practice this approach. Even the language of the recommendations for such treatment are more strongly worded; for example, the British Psychological Society (2004) stated, "Medication is sometimes a necessary intervention for ADHD though it is rarely sufficient alone" (p. 15). This is consistent, however, with the British attitude that the disorder is psychosocial in nature and not solely medical. The British solutions will therefore also be psychosocial and not solely medical.

This attitude has significant implications for educators. By focusing on behavioral approaches, education professions will necessarily play a large role in treating a child with ADHD. In the wording of one study, treating the disorder medically "disempowers" educators by ignoring the potential effect of altering the school environment. Utilizing a "functional approach" that recognizes the child's individual skills and environment factors "puts the power and responsibility for the intervention in the hands of educators" (Reid, Reason, Maag, Prosser, & Xu, 1998). Furthermore, the British attitude toward schooling, regardless of students' capacities, focuses on "environmental determinants of behavior," placing a great responsibility on the educator to ensure that students are engaged (Reason, 1999, p. 90). This attitude presumes that if children do not pay attention, the fault lies with the task of the adult responsible for the task.

In the British system, the responsibility of the educator is heavy and only becomes more so with the introduction of an ADHD student. The recommendations for British educators who exercise such power, however, are generally the same as for American and Australian educators. Techniques to be used include "positive reinforcement, token economies, contingency contracting, response cost, and time out" (Reid et al., 1998). Another source lists areas to address including the physical learning environment, classroom management, self-monitoring skills, and others

(Connor, Epting, Freeland, Halliwell, & Cameron, 1997). With such emphasis placed on the educator and relatively few innovative means of assistance, the implications for British educators can be serious. "Teachers are also more likely to experience a negative impact on their professional self-esteem" (Connor et al., 1997).

## CONCLUSION

After this discussion of ADHD—its definition, prevalence, and treatment in three different countries—it is plain to see that this is an issue which still requires a great deal of clarification. As we have seen, the criteria of the definition of ADHD play a large role in addressing all of the features of the illness. Without a clear definition, it will be impossible to achieve consistent or comparable rates of prevalence to establish how pervasive this illness really is. As a result, effective treatment strategies will be impossible to implement.

The discussion of ADHD and its effects on education also make it clear that a more unified and consistent approach is necessary to address the educational needs of these children. The consistency among the three countries studied in terms of classroom strategies, despite different attitudes toward the nature of the disorder, suggests that more work needs to be done to assist educators. A great burden is placed on teachers and other education professionals in dealing with children who show the symptoms of ADHD, and there should be more tactics and help available to those who remain responsible for these children's learning. Simplistic suggestions such as organizing the room formally and using both positive and negative reinforcement seem to be the same strategies already employed by teachers, regardless of inattention, impulsivity, or hyperactivity among students.

The United States and Australia are on a very similar path in terms of diagnosis, prevalence, and treatment of ADHD. The great variance on the part of the United Kingdom in these areas reveals an interesting attitude toward the illness and its constructs. The medical approach versus psychosocial approach debate that envelops ADHD is of course at the root of the variance. It would be beneficial for all countries if the research in this area were not so heavily dominated by the North American medical view. Further research may also reveal social and sociological roots to the debate. Great strides have been made in deciphering this illness, but in many ways, this progress has left many more questions.

## REFERENCES

- American Association of Pediatrics. (2001). *Clinical practice guideline: Treatment of the school-aged child with attention-deficit/hyperactivity disorder*. Retrieved March 24, 2007, from <http://aappolicy.aappublications.org/cgi/content/full/pediatrics;108/4/1033>



- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed., pp. 78–83). Washington, DC: American Psychiatric Press.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed, text revision). Washington, DC: American Psychiatric Press.
- Bramble, D. (2003). Annotation: The use of psychotropic medications in children: A British view. *Journal of Child Psychology and Psychiatry*, 44(2), 169–179. Retrieved from Medline database.
- British Psychological Society. (2004). *Attention deficit-hyperactivity disorder (AD-HD)—guidelines for multi-agency working*. Retrieved March 31, 2007 from [http://www.bps.org.uk/downloadfile.cfm?file\\_uuid=1B298FB9-7E96-C67F-DCDDFA0AF4A22811&ext=pdf](http://www.bps.org.uk/downloadfile.cfm?file_uuid=1B298FB9-7E96-C67F-DCDDFA0AF4A22811&ext=pdf)
- Ciechomski, L., Blashki, G., & Tonge, B. (2004). Common psychological disorders in childhood. *Australian Family Physician*, 33(12), 997–1005. Retrieved from Medline database.
- Concannon, P. E., & Tang, Y. P. (2005). Management of attention deficit hyperactivity disorder: a parental perspective. *Journal of Paediatrics and Child Health*, 41. Retrieved from Medline database.
- Connor, M. J., Epting, C., Freeland, R., Halliwell, S., & Cameron, R. J. (1997). Understanding and managing attention deficit hyperactivity disorder: The perspective from psychology. *Education and Child Psychology*, 14(1), 60–69. Retrieved from Medline database.
- DuPaul, G. J., & Weyandt, L. L. (2006). School-based intervention for children with AD/HD: effects on academic, social, and behavioural functioning. *International Journal of Disability, Development, and Education*, 53(2), 161–176. Retrieved from ERIC database.
- Faraone, S. V., Sergeant, J., Gillberg, C., & Biederman, J. (2003). The worldwide prevalence of ADHD: Is it an American condition? *World Psychiatry*, 2(2), 104–113. Retrieved from Medline database.
- Gomez, R., Harvey, C., Quick, L., Scharer, I., & Harris, G. (1999). DSM-IV AD/HD: Confirmatory factor models, prevalence, and gender and age differences based on parent and teacher ratings of Australian primary school children. *Journal of Child Psychology and Psychiatry*, 40(2), 265–274. Retrieved from Medline database.
- Isaacs, D. (2006). Attention-deficit/hyperactivity disorder: Are we medicating for social disadvantage? *Journal of Pediatrics and Child Health*, 42, 544–547. Retrieved from Medline database.
- Jick, H., Kaye, J., & Black, C. (2004). Incidence and prevalence of drug-tested attention deficit disorder among boys in the UK. *British Journal of General Practice*, 54, 345–347. Retrieved from Medline database.
- Kos, J. M., Richdale, A. L., & Hay, D. A. (2006). Children with attention deficit hyperactivity disorder and their teachers: a review of the literature. *International Journal of Disability, Development and Education*, 53(2), 147–160. Retrieved from ERIC database.
- National Health and Medical Research Council. (1997). *Attention deficit hyperactivity disorder*. Retrieved from <http://www.nhmrc.gov.au/publications/synopses/adhdsyn.htm>
- Parr, J. R., Ward, A., & Inman, S. (2003). Current practice in the management of attention deficit disorder with hyperactivity (ADHD). *Child: Care, Health & Development*, 29(3), 215–218. Retrieved from Medline database.
- Purdie, N., Hattie, J., & Carroll, A. (2002). A review of the research on intervention for attention deficit hyperactivity disorder: what works best? *Review of Educational Research*, 72(1), 61–99. Retrieved from JSTOR database.
- Reason, R. (1999). ADHD: A psychological response to an evolving concept (Report of a working party of the British Psychological Society). *Journal of Learning Disabilities*, 32(1), 85–91. Retrieved from Medline database.
- Reid, R., Reason, R., Maag, J., Prosser, B., & Xu, C. (1998). Attention deficit hyperactivity disorder in schools: a perspective on perspectives. *Educational and Child Psychology*, 15(4), 56–67. Retrieved from Medline database.
- Reid, R., & Maag, W. (1997). Attention deficit hyperactivity disorder: over here and over there. *Educational and Child Psychology*, 14(1), 10–20. Retrieved from Medline database.

Sawyer, M. G., Rey, J. M., Arney, F. M., Whitham, J. N., Clark, J. J., & Baghurst, P. A. (2004). Use of health and school-based services in Australia by young people with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, 43(11), 1355–1363. Retrieved from Medline database.

Sayal, K., Goodman, R., & Ford, T. (2006). Barriers to the identification of children with attention deficit/hyperactivity disorder. *Journal of Child Psychology and Psychiatry*, 47(7), 744–750. Retrieved from ERIC database.

Shaw, K., Wagner, I., Eastwood, H., & Mitchell, G. (2002). A qualitative study of Australian GPs' attitudes and practices in the diagnosis and management of attention-deficit/hyperactivity disorder (ADHD). *Family Practice*, 20, 129–134. Retrieved from Medline database.

Wolraich, M. L. (2003). Annotation: the use of psychotropic medications in children: An American view. *Journal of Child Psychology and Psychiatry*, 44(2), 159–168. Retrieved from ERIC database.

World Health Organization. (n.d.). *Mental and behavioural disorders*. Retrieved from International Classification of Diseases Web site: <http://www.who.int/classifications/apps/icd/icd10online/?gf90.htm+f90>

Zuvekas, S. H., Vitiello, B., & Norquist, G. S. (2006). Recent trends in stimulant medication use among U.S. Children. *American Journal of Psychiatry*, 163(4), 579–585. Retrieved from Medline database.

## APPENDIX A

### DSM–IV–TR CRITERIA FOR ADHD

#### A. Either (1) or (2)

- (1) 6 (or more) of the following symptoms of inattention have persisted for at least 6 months to a degree that is maladaptive and inconsistent with developmental level:

#### Inattention

- (a) often fails to give close attention to details or makes careless mistakes in schoolwork, work, or other activities
- (b) often has difficulty sustaining attention in tasks or play activities
- (c) often does not seem to listen when spoken to directly
- (d) often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (not due to oppositional behaviour or failure to understand instructions)
- (e) often has difficulty organising tasks and activities
- (f) often avoids, dislikes, or is reluctant to engage in tasks that require sustained mental effort (such as schoolwork or homework).
- (g) often loses things necessary for tasks or activities (e.g. toys, school assignments, pencils, books, or tools)
- (h) is often easily distracted by extraneous stimuli
- (i) is often forgetful in daily activities

- (2) 6 (or more) of the following symptoms of hyperactivity-impulsivity have persisted for at least 6 months to a degree that is maladaptive and inconsistent with developmental level

## Hyperactivity

- (a) often fidgets with hands or feet or squirms in seat
- (b) often leaves seat in classroom or in other situations in which remaining seated is expected
- (c) often runs about or climbs excessively in situations in which it is inappropriate (in adolescents or adults, may be limited to subjective feelings of restlessness)
- (d) often has difficulty playing or engaging in leisure activities quietly
- (e) is often “on the go” or often acts as if “driven by a motor”
- (f) often talks excessively

## Impulsivity

- (g) often blurts out answers before questions have been completed
- (h) often has difficulty awaiting turn
- (i) often interrupts or intrudes on others (e.g. butts into conversations or games)

- B. Some hyperactive-impulsive or inattentive symptoms that caused impairment were present before age 7 years.
- C. Some impairment from the symptoms is present in two or more settings (e.g. at school [or work] and at home).
- D. There must be clear evidence of clinically significant impairment in social, academic, or occupational functioning.
- E. The symptoms do not occur exclusively during the course of a Pervasive Developmental Disorder, Schizophrenia, or other Psychotic Disorder and are not better accounted for by another mental disorder (e.g. Mood Disorder, Anxiety Disorder, Dissociative Disorder, or a Personality Disorder)

314.01 ADHD, Combined Type – if both A1 and A2 for at least 6 months

314.00 ADHD, Predominantly Inattentive Type

314.01 ADHD, Predominantly Hyperactive-Impulsive Type

## APPENDIX B

### ICD-10 Criteria for Hyperkinetic Disorders (ADHD)

#### F90 Hyperkinetic disorders

##### G1 Inattention

At least six of the following symptoms of attention have persisted for at least six months, to a degree that is maladaptive and inconsistent with the developmental level of the child:

- (1) often fails to give close attention to details, or makes careless errors in school work, work or other activities;
- (2) often fails to sustain attention in tasks or play activities;
- (3) often appears not to listen to what is being said to him or her;
- (4) often fails to follow through on instructions or to finish school work, chores, or duties in the workplace (not because of oppositional behaviour or failure to understand instructions);
- (5) is often impaired in organising tasks and activities;
- (6) often avoids or strongly dislikes tasks, such as homework, that require sustained mental effort;
- (7) often loses things necessary for certain tasks and activities, such as school assignments, pencils, books, toys or tools;
- (8) is often easily distracted by external stimuli;
- (9) is often forgetful in the course of daily activities.

### G2 Hyperactivity

At least three of the following symptoms of hyperactivity have persisted for at least six months, to a degree that is maladaptive and inconsistent with the developmental level of the child:

- (1) often fidgets with hands or feet or squirms on seat;
- (2) leaves seat in classroom or in other situations in which remaining seated is expected;
- (3) often runs about or climbs excessively in situations in which it is inappropriate (in adolescents or adults, only feelings of restlessness may be present);
- (4) is often unduly noisy in playing or has difficulty in engaging quietly in leisure activities;
- (5) exhibits a persistent pattern of excessive motor activity that is not substantially modified by social context or demands.

### G3 Impulsivity

At least one of the following symptoms of impulsivity has persisted for at least six months, to a degree that is maladaptive and inconsistent with the developmental level of the child:

- (1) often blurts out answers before questions have been completed;
- (2) often fails to wait in lines or await turns in games or group situations;
- (3) often interrupts or intrudes on others (eg butts into others' conversations or games);
- (4) often talks excessively without appropriate response to social constraints.

G4 Onset of the disorder is no later than the age of seven years.

G5 Pervasiveness – The criteria should be met for more than a single situation, eg the combination of inattention and hyperactivity should be present both at home and at school, or at both school and another setting where children are observed, such as a clinic. (Evidence for cross-situationality will ordinarily require information from more than one source; parental reports about classroom behaviour, for instance, are unlikely to be sufficient.)



G6 The symptoms in G1 and G3 cause clinically significant distress or impairment in social, academic, or occupational functioning. G7 The disorder does not meet the criteria for pervasive developmental disorders (F84.-), manic episode (F30.-), depressive episode (F32.-), or anxiety disorders (F41.-).

Comment – Many authorities also recognise conditions that are sub-threshold for hyperkinetic disorder. Children who meet criteria in other ways but do not show abnormalities of hyperactivity/impulsiveness, may be recognised as showing attention deficit; conversely, children who fall short of criteria for attention problems but meet criteria in other respects may be recognised as showing activity disorder. In the same way, children who meet criteria for only one situation (eg only the home or only the classroom) may be regarded as showing a home-specific or classroom-specific disorder. These conditions are not yet included in the main classification because of insufficient empirical predictive validation, and because many children with sub-threshold disorders show other syndromes (such as Oppositional Defiant Disorder, F91.3) and should be classified in the appropriate category.

### **F90.0 Disturbance of activity and attention**

The general criteria for hyperkinetic disorder (F90) must be met, but not those for conduct disorders (F91.-).

### **F90.1 Hyperkinetic Conduct Disorder**

The general criteria for both hyperkinetic disorder (F90) and conduct disorders (F91.-) must be met.

### **F90.8 Other hyperkinetic disorder**

### **F90.9 Hyperkinetic disorder, unspecified**

This residual category is not recommended and should be used only when there is a lack of differentiation between F90.0 and F90.1 but the overall criteria for F90.- are fulfilled.









## INSTRUCTIONS FOR AUTHORS

**Content:** *Peabody Journal of Education (PJE)* publishes quarterly symposia in the broad area of education, including but not limited to topics related to formal institutions serving students in early childhood, preschool, primary, elementary, intermediate, secondary, postsecondary, and tertiary education. The scope of the journal includes special kinds of educational institutions, such as those providing vocational training or the schooling for students with disabilities. PJE also welcomes immediate framework of institutions, and education matters that are important to nations outside the United States. Finally, it includes topics that are linked to the social and organizational context in which formal and informal education take place. The Editor cooperates with groups of scholars to present multifaceted, integrated expositions of important topics. A given issue of PJE may contain contributions from social scientists, historians, philosophers, attorneys, practitioners, and policymakers. Unsolicited proposals for special issues - including designation of participating scholars and an outline of articles - will be accepted for review. Additionally, the Editor cooperates with Editorial Board members to identify potential topics, guest editors, and contributors. PJE has the flexibility to consider publishing monographs or a series focused on particular lines of inquiry. In all cases, the Editor and the Editorial Board will ensure that each issue is carefully reviewed and its articles will comprise a high-quality contribution to understanding and practice.

**Manuscript Preparation:** To submit a prospectus for a special issue, note submission requirements presented here and contact the Coordinating Editor, Michele Thompson, for complete details (address below). Prepare manuscripts according to the *Publication Manual of the American Psychological Association* (5th ed., 2001; available from APA Order Department, P.O. Box 2710, Hyattsville, MD 20874), especially with regard to reference lists and text citations. Follow Guidelines to Reduce Bias in Language (*APA Manual*, pp. 61-87). Manuscripts should include the theme title and rationale for the theme. Using 8.5 x 11-in., nonsmear paper and 1-in. margins, type all components double spaced and in the following order: title page (p. 1), abstract (p. 2), text (including quotations), references, appendices, footnotes, tables, and figure captions. On page 1, type article title, author name(s) and affiliation(s), running head (abbreviated title, no more than 45 characters and spaces), author notes and acknowledgments, and name and address of person to whom requests for reprints should be addressed; on page 2, type an abstract of no more than 150 to 200 words. Indent all paragraphs. Use footnotes sparingly. Attach photocopies of all figures. Number all manuscript pages (including photocopies of figures).

**Cover Letter, Permissions, Credit Lines:** In a cover letter, include the contact author's complete mailing address, e-mail address, and telephone and fax numbers. State that the manuscript includes only original material that has not been published and that is not under review for publication elsewhere. Authors are responsible for all statements made in their work and for obtaining permission to reprint or adapt a copyrighted table or figure or to quote at length from a copyrighted work. Authors should write to original author(s) and original publisher to see if permission is required and to request nonexclusive world rights in all languages to use the material in the current article and in future editions. Include copies of all permissions and credit lines with the manuscript. (See p. 175 of the *APA Manual* for samples of credit lines.)

**Manuscript and Computer Disk Submissions:** Prepare manuscript on a word processor and submit three (3) high-quality printouts and a disk to the Coordinating Editor, Michele Thompson, *Peabody Journal of Education*, Department of Leadership, Policy, and Organizations, Peabody College of Vanderbilt University, Peabody #41, 230 Appleton Place, Nashville, TN 37203-5721. The disk should contain two files - word processor and ASCII versions of the manuscript - and file content must match the printed manuscript exactly, or there will be a delay in publication. Manuscripts and disks are not returned. **Accepted Manuscripts:** Authors are strongly encouraged to submit manuscripts on disk. The disk should be prepared using MS Word or WordPerfect and should be clearly labeled with the authors' names, file name, and software program. A hardcopy printout that exactly matches the disk must be supplied. Each manuscript must be accompanied by a statement that it has not been published elsewhere and that it has not been submitted simultaneously for publication elsewhere. Authors are responsible for obtaining permission to reproduce copyrighted material from other sources and are required to sign an agreement for the transfer of copyright to the publisher. All accepted manuscripts, artwork, and photographs become the property of the publisher.

All parts of the manuscript should be typewritten, double-spaced, with margins of at least one inch on all sides. Number manuscript pages consecutively throughout the paper. Authors should also supply a shortened version of the title suitable for the running head, not exceeding 50 character spaces. Each article should be summarized in an abstract of not more than 100 words. Avoid abbreviations, diagrams, and reference to the text in the abstract.

**Illustrations:** Illustrations submitted (line drawings, halftones, photos, photomicrographs, etc.) should be clean originals or digital files. Digital files are recommended for highest quality reproduction and should follow these guidelines:

- 300 dpi or higher
- Sized to fit on journal page
- EPS, TIFF, or PSD format only
- Submitted as separate files, not embedded in text files

Color illustrations will be considered for publication; however, the author will be required to bear the full cost involved in their printing and publication. The charge for the first page with color is \$900.00. The next three pages with color are \$450.00 each. A custom quote will be provided for color art totaling more than 4 journal pages. Good-quality color prints should be provided in their final size. The publisher has the right to refuse publication of color prints deemed unacceptable.

**Tables and Figures.** Tables and figures (illustrations) should not be embedded in the text, but should be included as separate sheets or files. A short descriptive title should appear above each table with a clear legend and any footnotes suitably identified below. All units must be included. Figures should be completely labeled, taking into account necessary size reduction. Captions should be typed, double-spaced, on a separate sheet. All original figures should be clearly marked in pencil on the reverse side with the number, author's name, and top edge indicated.

**Proofs and Reprints.** Page proofs are sent to the designated author using Taylor & Francis' Central Article Tracking System (CATS). They must be carefully checked and returned within 48 hours of receipt. Reprints of individual articles are available for order at the time authors review page proofs. A discount on reprints is available to authors who order before print publication.

# Peabody Journal of Education

Volume 83, Number 1, 2008

Introduction to the Special International Issue <i>Stephen P. Heyneman</i>	1
The Origins of the International Standard Classification of Education <i>John A. Smyth</i>	5
Implementing Free Primary Education Policy in Malawi and Ghana: Equity and Efficiency Analysis <i>Kazuma Inoue and Moses Oketch</i>	41
Globalization and Implementation of an Equity Norm in Higher Education: Admission Processes and Funding Framework Under Scrutiny <i>Gaële Goastellec</i>	71
Aspects of Fiscal Federalism in Higher Education Cost Sharing in Latvia <i>Rita Kasa</i>	86
Higher Educational Cost-Sharing, Dual-Track Tuition Fees, and Higher Educational Access: The East African Experience <i>Pamela Marcucci, D. Bruce Johnstone, and Mary Ngolovo</i>	101
Do Electronic Technologies Increase or Narrow Differences in Higher Education Quality Between Low- and High-Income Countries? <i>Norman Clark Capshaw</i>	117
Compulsion, Craft, or Commodity? Education Services Trade in the Larger Context <i>Brandyn L. Payne</i>	133
Diagnosis, Treatment, and Educational Implications for Students With Attention-Deficit/Hyperactivity Disorder in the United States, Australia, and the United Kingdom <i>Sarah Schlachter</i>	154



0161956X(2008)83(1)

ISBN 080589250-8



9 790805 892504

9000



Volume 83 Issue 2 ISSN 0161-956X 2008

# *Peabody Journal of Education*

**CURRENT YR/VOL**

**Marygrove College Library  
8425 West McNichols Road  
Detroit, MI 48221**

published in cooperation with the Peabody College of Vanderbilt University

ROUTLEDGE



**Routledge**  
Taylor & Francis Group



# PEABODY JOURNAL OF EDUCATION

Camilla P. Benbow, Dean, Peabody College  
James W. Guthrie, Editor  
Michele Thompson, Coordinating Editor  
Deborah M. Enright, Associate Editor

## EDITORIAL ADVISORY BOARD

Darlyne Bailey, *University of Minnesota*  
R. Dale Ballou, *Peabody College of Vanderbilt University*  
Sir Clive Booth, *Oxford Brookes University, England*  
Mary M. Brabeck, *New York University*  
Mark Bray, *The University of Hong Kong, PRC*  
Louis A. Castenell, Jr., *The University of Georgia*  
David Y. Chen, *Chongquin Technology and Business University (CTBU)*  
Mike Dishman, *McLocklin, Murphy & Dishman, L.L.P.*  
Lora Cohen-Vogel, *Florida State University*  
Joseph P. Du Cette, *Temple University*  
Victor Firsov, *Russian Academy of Teachers Retraining, Russia*  
Juan Froemel, *LLECE/UNESCO, Chile*  
Louis M. Gomez, *Northwestern University*  
Gerardo M. Gonzalez, *Indiana University*

David W. Gordon, *Elk Grove Unified School District, California*  
Eric A. Hanushek, *Hoover Institution, Stanford University*  
Stephen P. Heyneman, *Peabody College of Vanderbilt University*  
Phillip W. Jones, *The University of Sydney, Australia*  
Kazuo Kuroda, *Waseda University, Japan*  
James H. Lytle, *Trenton Public Schools, New Jersey*  
Michael K. McLendon, *Peabody College of Vanderbilt University*  
Ann L. Marcus, *New York University*  
Bill McKinney, *Region IV Education Service Center, Texas*  
Francois Orivel, *University of Bourgogne, France*  
Penelope L. Peterson, *Northwestern University*  
Michael J. Podgursky, *University of Missouri-Columbia*  
Robert Slavin, *Johns Hopkins University*  
Jandhyala B. G. Tilak, *National Institute of Education Planning and Administration, India*

---

**Abstracted and/or Indexed in:** PsycINFO/Psychological Abstracts; Linguistics and Language Behavior Abstracts; Contents Pages in Education; Sociological Abstracts; Education Index; Education Abstracts; Social Services Abstracts; EBSCOhost Products; Cabell's Directories; Family Index Database.

*Peabody Journal of Education* (ISSN: 0161-956X) is published quarterly in February, May, August, and November for a total of 4 issues per year by Taylor and Francis Group, LLC, 325 Chestnut Street, Suite 800, Philadelphia, PA 19106.

**US Postmaster:** Please send address changes to *Peabody Journal of Education*, Taylor & Francis Group, LLC, 325 Chestnut Street, Suite 800, Philadelphia, PA 19106.

### Annual Subscription, Volume 83, 2008

Print ISSN - 0161-956X, Online ISSN - 1532-7930

Institutional subscribers: US \$462, UK £277, €370

Personal subscribers: US \$58, UK £35, €46

Institutional and individual subscriptions include access to the online version of the journal. Institutional subscriptions include access for any number of concurrent users across a local area network. Individual subscriptions are single username/password only.

**Production and Advertising Office:** 325 Chestnut Street, Suite 800, Philadelphia, PA 19106. Tel: 215-625-8900, Fax: 215-625-8563.  
Production Editor: Michelle McCarthy

### Subscription offices

**USA/North America:** Taylor and Francis Group, LLC, 325 Chestnut Street, Suite 800, Philadelphia, PA 19106. Tel: 215-625-8900, Fax: 215-625-2940.

**UK/Europe:** Taylor and Francis Customer Services, Sheepen Place, Colchester, Essex CO3 3LP, UK. Tel: +44 (0)-20-7017-5544, Fax: +44 (0)-20-7017-5198.

For a complete guide to Taylor and Francis Group, LLC's journal and book publishing programs, visit our website: [www.taylorandfrancis.com](http://www.taylorandfrancis.com)

**Copyright © 2008 Taylor and Francis Group, LLC.** All rights reserved. No part of this publication may be reproduced, stored, transmitted, or disseminated in any form or by any means without prior written permission from Taylor and Francis Group, LLC. Taylor and Francis Group, LLC grants authorization for individuals to photocopy copyright material for private research use. The publisher assumes no responsibility for any statements of fact or opinion expressed in the published papers. The appearance of advertising in this journal does not constitute an endorsement or approval by the publisher, the editor, or the editorial board of the quality or value of the product advertised or of the claims made for it by its manufacturer.

**Permissions.** For further information, please visit: <http://www.tandf.co.uk/journals/permissions.html>

May 2008



# Introduction to the Special Issue on From Equity to Adequacy to Choice

Michael Podgursky  
*University of Missouri–Columbia*

On October 30, 2007, the Truman School of Public Affairs at the University of Missouri and the Missouri Show-Me Institute hosted a 1-day conference on school finance titled “From Equity to Adequacy to Choice.” The motivation for this conference was a recent school finance “adequacy” lawsuit in the state and the ongoing problems of our large urban school districts. The largest district in the state (St. Louis) is now unaccredited and under state receivership, yet spending per student in the St. Louis district is nearly the highest in the state. The second major urban district—Kansas City—is likely to lose accreditation in the near future.

This situation is hardly unique to Missouri. Many states have ongoing school finance litigation. By one count, there have been 125 school cases challenging the constitutionality of state school finance systems, and 23 states have had their state funding systems ruled unconstitutional on adequacy grounds (Guthrie & Springer, 2007). Similarly, many of the states that have seen the most significant cases—New Jersey, for instance—also have major urban districts combining very high spending per student with abysmally low measures of student performance.

It was this dilemma that motivated this conference. We wanted to invite some of the leading scholars and litigators in the areas of school finance and school choice to write articles reflecting on aspects of this school finance and school failure conundrum. In particular, we were interested in new trends in this school finance and choice-related litigation. Was there common ground? The result is a collection of thoughtful yet provocative articles exploring a variety of areas on this topic.

The first piece, by Podgursky, Smith, and Springer, reviews the recent Missouri case that motivated the conference. As the title of their article suggests, the

Missouri case was highly unusual in that a group of choice-oriented taxpayers were given intervenor status for the defense in this school finance case. In this study the authors review some of the issues that arose in this and similar school finance cases. They point out the trade-off between horizontal and vertical equity. On the question of “adequacy,” they show that it is simply impossible to establish statistically any reliable relationship between the level of school spending and student performance on state assessments.

This theme is taken up and expanded by Costrell, Hanushek and Loeb in the second article. In light of the extensive litigation, legislatures and courts have been seeking scientific, or at least defensible, estimates of how much spending per student is required to produce certain levels of achievement, variously measured. Many approaches are offered in trials (a critical review of these may be found in Hanushek, 2006). However, one approach—the “cost function approach”—has begun to appear in school finance cases and in situations where state legislatures want estimates of necessary spending. Indeed, two different expert witnesses used this approach in the Missouri case. On the face of it, the cost function approach seems to be well grounded in economics—and microeconomics, in particular. However, these authors show that these estimates are not “costs” at all, but rather district spending levels. Their careful dissection shows that “adequacy” estimates based on such approaches have no statistical validity or reliability.

Jay Green and Julie Trivitt take up an interesting, but overlooked, question. What has this litigation done for student achievement? They analyze panel data on state National Assessment of Educational Progress (NAEP) test scores and search for breaks in trends associated with successful school finance adequacy cases. Given the statistically tenuous relationship between test scores and spending, it would be surprising to find a relationship between test scores and successful litigation. In fact, no such surprises are forthcoming. After careful analysis, they find no evidence in NAEP test data that plaintiff victories in “adequacy” lawsuits raise relative state test scores, reduce intrastate inequality, or raise state graduation rates.

Given the tenuous relationship between spending and achievement, Paul Hill addresses the challenging question for courts and education policymakers in “Spending Money When It Is Not Clear What Works.” He begins by carefully establishing that for minority and poor children, this is not a hyperbole. However, his is not a call for throwing in the towel. Rather he argues for a strong research and development (R&D)–based system that will help identify promising practices and diffuse them more rapidly to schools, using charter schools as R&D demonstration sites.

James Guthrie offers a sweeping, critical, and admittedly “polemical” overview of the development of the school finance litigation and policy debate. He provides an interesting discussion of successful reforms in American education and argues that in all cases, including the early rounds of school finance equity litigation,

they were initiated externally by well-organized, typically bipartisan reformers. He then offers some thoughts on how a similar coalition might be forged in the next stage of school finance reform.

The next two articles are by lawyers who have been involved in school choice litigation. Both take up the question of introducing choice into the menu of remedies in school finance cases. Clint Bolick, unquestionably the most experienced litigator in the United States in the area of school choice, reviews key developments in school finance litigation and sees opportunities to introduce choice as a remedy. Adequacy decisions, he argues, have opened the door for school choice plaintiffs to use similar arguments for allowing state funds to travel with the child. He takes the long view and compares the developments in school finance litigation with the long gestation period in school finance and civil rights cases. If school choice litigation is a long campaign, Julio Gomez describes in detail an opening skirmish—*Crawford v. Davy*. The New Jersey school finance case—*Robinson v. Cahill*—was one of the most important school finance cases in the United States and was an important victory for school finance equity advocates. This case laid the basis for the current regime in which 31 plaintiff districts (so-called Abbott districts) have received a huge infusion of state funds as a result of the case. Several of these districts now spend nearly \$20,000 per student. Yet student achievement in the Abbott districts is still very low. In light of this experience, school choice advocates attempted to use Abbott-like arguments for the court to open up a school choice remedy. Gomez provides a detailed analysis of this important case.

In short, fairness and efficacy in state school finance systems is a moving target. The 35 years since the *Serrano* decision in California have witnessed huge upheavals in the way K-12 education is financed. Unfortunately, achievement gaps have not been closed, or even narrowed to any significant degree. The contributors to this volume offer views that can help break the cycle of litigation and policy deadlock that characterizes public school finance today, potentially in ways that can serve taxpayers, parents, and children better.

## REFERENCES

- Guthrie, J. W., & Springer, M. G. (2007). Alchemy: Adequacy advocates turn guesstimates into gold. *Education Next*, 7(1), 20–27.
- Hanushek, R. (2006). Science violated: Spending projections and the “costing out” of an adequate education. In E. A. Hanushek (Ed.), *Courting failure: How school finance lawsuits exploit judges' good intentions and harm our children*. (pp. 257–310) Stanford, CA: Hoover Institution.

# A New Defendant at the Table: An Overview of Missouri School Finance and Recent Litigation

Michael Podgursky

*University of Missouri–Columbia*

James Smith

*Management Analysis and Planning*

Matthew G. Springer

*Department of Leadership, Policy, and Organizations Peabody  
College of Vanderbilt University*

Like many other states, Missouri has gone through several rounds of school finance litigation. However, the trial just concluded was unusual in two respects. First, three taxpayers were allowed to intervene for the defense and, in the process, raise important questions concerning the efficiency of school spending and broader questions of school reform. Second, the outcome at the circuit court level, which focused nearly entirely on points of law, was a complete victory for the defense. This article provides an overview of disputes of Missouri school finance and evidence pertaining to some of the points in dispute at the trial. These lessons generalize to other states facing school finance litigation. The authors conclude that changes in school funding formulas, and the seemingly interminable litigation about those formulas, are not an effective vehicle for addressing achievement gaps or the overall level of school performance.

On October 17, 2007, a Missouri circuit court handed down a major decision upholding the constitutionality of the current Missouri school finance system.

---

This article was prepared for the “From Equity to Adequacy to Choice” Conference, Truman School of Public Affairs/Missouri Show-Me Institute, October 30, 2007. The usual disclaimers apply.

Correspondence should be sent to Michael Podgursky, Department of Economics, University of Missouri, Columbia, MO 65211. E-mail: PodgurskyM@missouri.edu



However, this is but one milestone in a long road of litigation stretching back more than a decade. In this regard, Missouri is hardly unusual. Many other states have experienced prolonged litigation surrounding school finance. What made the Missouri case unusual—indeed, unique—was the fact that a group of three taxpayers intervened for the defense. This not only raised the overall vigor and quality of the defense but also provided a vehicle for raising questions about efficiency with which schools use their current funds, and it opened the door, at least a crack, for testimony about market-based school reforms and value-added measures as alternative remedies to the complaints of the plaintiffs.

The plaintiffs in the Missouri case are three groups of school districts, 264 in all, representing roughly 60% of Missouri public school enrollments. Each group of districts has a bill of grievances against the current school finance regime. The nominal defendants are various state officials and the Missouri Board of Education, but in practice the real defendant was the state legislature, which crafted the school finance law.

This article provides a survey of some of the key issues of education finance considered in this case. We begin with a general background on school finance litigation nationally. This is followed by some specifics of the Missouri case, including the historic intervention by taxpayer defendants. We then consider evidence in three key areas: the overall level and trend in K-12 resources in Missouri, ways of measuring the fairness (equity) in the distribution of these resources, and finally the relationship between these education resources and student achievement. We close with some observations on the efficacy of using school funding formulas to address student achievement gaps.

## BACKGROUND

### *National Litigation in School Finance*

Equity and adequacy are perhaps the two most prominent principles in school finance policy. Broadly speaking, school finance equity refers to fairness in the distribution of educational goods and services. Adequacy is less well defined. However, to proponents it usually refers to the availability of a sufficient level of resources for all students to reach some level of performance. Often the latter is defined in reference to performance on state assessments (e.g., “proficient”), although some courts have talked about more nebulous goals. For example, the Kentucky Supreme Court in *Rose v. Council for Better Education* laid out seven capacities that must be the goal for every child under a constitutionally “efficient” system of education:

1. Sufficient oral and written communication skills to enable students to function in a complex and rapidly changing civilization.
2. Sufficient knowledge of economic, social, and political systems to enable the student to make informed choices.
3. Sufficient understanding of governmental processes to enable the student to understand the issues that affect his or her community, state, and nation.
4. Sufficient self-knowledge and knowledge of his or her mental and physical wellness.
5. Sufficient grounding in the arts to enable each student to appreciate his or her cultural and historical heritage.
6. Sufficient training or preparation for advanced training in either academic or vocational fields so as to enable each child to choose and pursue life work intelligently.
7. Sufficient levels of academic or vocational skills to enable public school students to compete favorably with their counterparts in surrounding states, in academics or in the job market.

School finance litigation has gone through three broad phases, shifting from a focus on the distribution of educational resources to attempts to establish a relationship among education inputs, processes, and outcomes (Guthrie, Springer, Rolle, & Houck, 2007). The first phase ran from the late 1960s until 1973 and was adjudicated under the U.S. Constitution's equal protection clause. The second phase began with the U.S. Supreme Court's 5–4 decision in *San Antonio Independent School District v. Rodriguez* 411, U.S. 1 (1973), which concluded, in part, that education is not a “fundamental right” under the U.S. Constitution's equal protection clause. As a consequence, and for nearly 2 decades, school finance litigation relied on state constitutions' equal protection education clauses to guide legal challenges against state funding structures. The third phase of school finance litigation started when the Kentucky Supreme Court declared the state's entire system of public and elementary and secondary education unconstitutional and held that all Kentucky schoolchildren had a constitutional right to an adequate educational opportunity (*Rose v. Council for Better Education*, 1989).

The number of legal challenges against school funding mechanisms is quite substantial. More than 125 court cases challenging the constitutionality of school district and school spending levels have been filed since the late 1960s, an average of slightly more than 3 cases per year (Guthrie & Springer, 2007). Of these challenges, 12 states have had their state funding mechanism ruled unconstitutional on equity grounds and 23 states have had their state funding mechanism ruled unconstitutional on adequacy grounds. Only 5 states—Delaware, Hawaii, Mississippi, Nevada, and Utah—have not had their state school funding mechanisms adjudicated in the courts.

The cumulative impact of school finance litigation on school spending is considerable. In a recent state-by-state measure of the long-term fiscal impact of court mandated school finance reform, Atkins (2007) estimates that lawmakers have authorized an additional \$34 billion in annual spending or taxes to comply with court mandated reform since 1977. Although Atkins noted that the majority of these states (18) either have spent surplus funds or cut spending in other program areas to meet court directives, 9 states have raised taxes by a total of almost \$13 billion per annum. Atkins's estimates do not take into consideration the tens of millions of dollars incurred by taxpayers in school finance litigation.

The Missouri case is unusual in that, unlike any other such case, a substantial portion of the defense was borne by private citizens. Litigation is expensive. It is difficult to determine total costs of litigation in school finance cases because in most instances parties are not required to report this information. Sunshine Act requests have established that the plaintiff districts in Missouri have spent roughly \$3.2 million since 2004 on this case—and this does not count the litigation costs for the defense.<sup>1</sup> When the Wyoming legislature was facing its fourth round of litigation in *Campbell County v. Wyoming*, it finally required the plaintiff school districts to report all litigation-related expenses. They reported spending \$2,886,122 on the fourth trial alone. The state's counsel estimates that his agency spent about half that amount on the same trial. Likewise, in a recent South Carolina case, plaintiffs' attorneys reported fees and costs of approximately \$6 million, and defense attorneys estimated fees and costs of approximately \$3.5 million.<sup>2</sup> In both instances the amounts reported include only monetary expenses and not the opportunity cost of the time and effort used by school district and state agency personnel in responding to requests for information, being deposed, testifying, and so on.<sup>3</sup>

School finance litigation has also shaped state school funding mechanisms. In 1998, for example, Murray, Evans, and Schwab (1998) concluded that as a result of court-mandated reform, intrastate inequality was dampened to the point that disparities between states were greater than disparities within states. The authors also concluded that spending rose in the lowest and median spending school districts and remained constant in the highest spending school districts.

---

<sup>1</sup>Taxpayer-funded spending to date totals at least \$5.0 million. This includes \$3.6 million by the school districts and \$1.4 million by the state Attorney General's Office (AGO) to a private law firm for assistance in the litigation. (Franck, 2007). These totals do not include the time of the AGO staff.

<sup>2</sup>Communication: August 2007. *Abbeville County School District v. State*, 515 S.E. 2d 535 (S.C. 1999) (state's attorney and counsel for the state legislature, personal communication).

<sup>3</sup>It is likely that plaintiffs view expenditures for litigation as cost effective. First, and perhaps foremost, these costs are borne by taxpayers, not the individuals involved. Second, winning almost always results in significant increases in school district revenues. Finally, increases in school district revenues almost always result in pay raises for teachers and administrators (see, e.g., Clark, 2003; Hanushek & Rivkin, 1997).

A more recent study by Springer, Liu, and Guthrie (2007), however, examined whether differences in resource allocation patterns exist between equity- versus adequacy-based reform. They found that both equity- and adequacy-based school finance reform resulted in changes in a state's funding mechanism. However, the authors neither detected any differences between court-mandated equity- versus adequacy-based reform nor discovered any evidence of adequacy-based reform resulting in the allocation of additional resources to low wealth districts when compared to outcomes under court-mandated equity reform. Berry (2007) undertook a similar examination, making some statistical corrections for serial correlation and found generally very weak effects of either type of litigation on a range of fiscal variables.

### *Missouri Litigation*

Missouri's school finance system was challenged on equity grounds and found unconstitutional in 1993.<sup>4</sup> The legislature responded by passing the School Improvement Act of 1993, which called for an extensive overhaul of the school funding mechanism by means of an increase in elementary and secondary education spending and decoupling local tax collections from local wealth. The legislature put in place a financing formula meant to reduce the link between district wealth per student and district school revenues. In theory, if a property-rich and a property-poor district had the same tax rate, the state would make up the difference and equalize revenues. If a school district exerted the appropriate tax effort, it was guaranteed the tax revenues of a school district at the 95th percentile of property wealth. Thus, a school district with one fourth the wealth of a 95th percentile district would get \$3 of state aid for every \$1 of local revenue.

This was an ambitious goal and the state legislature was never able to fully meet this revenue target. Although real spending per student rose briskly in the years following this reform (see Figure 2), the legislature slipped further behind in meeting the SB380 spending target. A recession in 2001 led to cuts in almost all areas of state spending except K-12 education. However, although real spending on K-12 education rose, it did not rise fast enough to meet this funding target. The problem is that full funding under this formula required that state spending track property values in the wealthiest (95th percentile) districts. Although state income rose over this period, it did not keep up with the rise in housing prices, particularly in the wealthiest districts (Podgursky & Springer, 2006).

---

<sup>4</sup>*Committee for Educational Equality v. State of Missouri* (1993)



Restive school districts threatened another lawsuit if the SB380 formula was not “fully funded.” In 2004, in preparation for an “adequacy” lawsuit, the Missouri School Boards Association contracted with a consulting firm (Augenblick & Meyers, 2003) to conduct an “adequacy” study of Missouri school spending.<sup>5</sup> These consultants approached this question in two ways. The first, a “professional judgment” approach, was to put together panels of educators and administrators to ask what level of spending would be required to meet student achievement goals in a school. The second, sometimes labeled the “successful schools” approach, was to examine the spending levels of school districts that met Department of Elementary and Secondary Education (DESE) performance targets. Both approaches found large spending shortfalls. Augenblick and Myers reconciled the recommendations of the two approaches and concluded that in 2001–02 Missouri was underspending by \$913 million, not including cost of transportation, food services or capital expenditures. Myers testified at trial that even at increased spending in SB287, Missouri was underspending by \$800 million.

In response to these concerns, and in an attempt to forestall the adequacy lawsuit, the legislature adopted a “successful schools” approach in the new funding scheme. The latter approach formed the basis for the new spending scheme adopted by the legislature in 2005. The legislature took as their target guaranteeing that every school district in the state would have revenues at least that of 113 school districts designated “distinguished” in the 2003–04 academic year by DESE. The “distinguished” designation was computed on the basis of the level or gains in student achievement. The new law, SB 287, represented an overhaul of the state funding formula.

In 2005, the legislature determined that the minimum adequate level of spending was \$6,117 per student. The legislature arrived at this figure by calculating the average operating spending per student for the 113 districts with perfect or nearly perfect scores on the Annual Performance Report conducted by DESE. Annual Performance Report scores are heavily weighted toward performance on the Measures of Academic Progress assessment. This figure will be recomputed every 2 years. In theory, the figure could go down. However, SB287 specifies that the old level will stay in effect should that occur.

Although the legislature had hoped to forestall an “adequacy” lawsuit by adopting a successful schools approach, a group of plaintiff’s districts chose to proceed with litigation. The lawsuit, originally filed in 2004, was reactivated and went to trial in January through March 2007.

---

<sup>5</sup>See trial testimony of Robert Costrell for a critique of the methodology Augenblick and Myers employed in their “adequacy” study.

### *A New Defendant at the Table*

Until this point, the Missouri case resembled dozens previously filed in many other states. On the plaintiff side was a core group of 236 heavily rural or small-town districts (Committee for Educational Equality). These districts typically, although not consistently, had below-average spending per student. They were joined by two other interveners on the plaintiff's side. The first was the Coalition to Fund Excellent Schools, a group of 28 wealthy suburban school districts. The second was the St. Louis school district.

The plaintiffs and interveners thus included the wealthiest and poorest school districts in the state, with a broad group in between. What explains this odd collection? In fact, this type of strategy is not uncommon in school finance cases. The wealthy Coalition to Fund Excellent Schools districts, in particular, wanted to make sure that whatever the court might decide would not come at their expense. In a case such as this, one might easily imagine remedies that would redistribute state aid from wealthier districts to the poorer ones. Intervening as co-plaintiffs was a way for the wealthy districts to steer the case away from such threatening territory. The St. Louis district's participation was motivated by its unique struggles—indeed, because the trial it has lost accreditation and has effectively been placed in receivership by the state. As one of the highest spending districts in the state, their goal was to avoid any remedy that might come at their expense.

The named defendants in this case included the state legislature, DESE, and various other government officials. In practice, the defendant was the state legislature, as it was the architect of the school finance system in dispute. As in all other such cases, it is the duty of the Attorney General (AG) and his staff to defend the state.<sup>6</sup> Thus, in the courtroom the “defense” team consists of lawyers and staff from the AG's office, and the plaintiffs are lawyers and staff from law firms representing the three plaintiff groups.

What is unusual, and in fact historic, about this case is that three taxpayers intervened as defendants. In the period leading up to the trial there was concern in some quarters that the AG, a Democrat with clear gubernatorial ambitions (the governor is Republican), was not preparing an adequate defense. In a petition to the court, the three taxpayer interveners (Rex Sinquefield, Menlo Smith, and Bevis Schock) claimed that the defense was doing an “incompetent” job. There was a good deal of evidence that the defense was poorly prepared as the trial approached.<sup>7</sup> Although the court rejected their claim of incompetence, they did admit them as defendants. We are aware of no other school finance case in the country where this has happened.

---

<sup>6</sup>In several states the AG, or other state agency, contracts with a private law firm to represent the state.

<sup>7</sup>See Intervenors' Motion to Intervene.

The presence of these interveners changed the dynamic of the case. Financed entirely by private donations, they hired an aggressive and experienced trial lawyer. They also brought in several experts from around the country with extensive school finance trial experience. More important, they acted independently to raise questions concerning school efficiency and remedies such as school choice, that were not part of the AG's defense. Thus, the judge heard a much wider range of opinions and options than would have otherwise been the case.

## SELECTED POINTS OF CONTENTION

As is common practice in adequacy cases, the plaintiffs confronted the court with a barrage of concerns and complaints. This tactic may arise because a diverse group of plaintiff districts (who are paying legal fees) want their particular grievances aired, or merely a hope on the plaintiff's part that some complaint out of the potpourri will resonate with the judge.

In this case the list of complaints was replete with issues and charges about nearly every conceivable aspect of school funding in Missouri. Of course, the leitmotif of the case was that the state did not provide school districts enough money, but also the claim that the Missouri school finance formula resulted in inequitable funding among school districts occasioned multiple expert reports and considerable testimony from several witnesses. We describe this issue in more detail later, but first we consider the list of other complaints raised by Plaintiffs. These complaints are most interesting primarily because of their breadth and diversity. They are described approximately in the order they were addressed in Plaintiffs' "Finding of Facts" submitted to the court at the conclusion of the trial.

Plaintiffs' global charge was that the state "violated the Missouri Constitution through disparities, inadequacies and inequalities of the school funding formula . . . new, increased and expanded requirements have been funded in violation of the Hancock Amendment to the Missouri Constitution . . . [and] the funding for these requirements has been shifted to . . . districts and to the taxpayers."<sup>8</sup>

---

<sup>8</sup>Article X, Section 16, of the Missouri Constitution provides in part as follows: "The state is prohibited from requiring any new or expanded activities by counties and other political subdivisions without full state financing, or from shifting the tax burden to counties and other political subdivisions." Article X, Section 21, of the Missouri Constitution provides as follows: "The state is hereby prohibited from reducing the state financed portion of the costs of any existing activity or service required of counties and other political subdivisions. A new activity or service or an increase in the level of any activity or service beyond that required by existing law shall not be required by the general assembly or any state agency of counties or other political subdivisions, unless a state appropriation is made and disbursed to pay the county or other political subdivision for any increased costs." Plaintiff CEE (Committee for Educational Equality) Findings of Facts, May 2007.

The plaintiffs argued that the new funding formula either failed to provide funding for students identified as gifted or failed to provide sufficient funding for such students. They charged that the state did not provide funding for transportation, employee background checks, compliance with federal Individuals with Disabilities Act, or homeless student services. They further complained that the state did not adequately fund the costs of state-mandated graduation requirements, student assessments, or early childhood education programs. They argued that the 7-year phase-in of the new school finance formula was unconstitutional, that the state did not provide adequate funding to build and maintain facilities, and that the regional cost adjustment (Dollar Value Modifier) failed to compensate for the higher costs of urban districts. Several witnesses testified that the weights the legislature used to calculate the costs for special needs students (limited English proficient [LEP], economically disadvantaged, and special education) were inadequate.<sup>9</sup>

### *The Level and Distribution of District Spending Per Student*

Assessment of the plaintiffs' "adequacy" claims must begin with the Missouri constitution. As in many other states, the Missouri constitution provides for free public schools. Section 1(a) of the constitution states:

A general diffusion of knowledge and intelligence being essential to the preservation of rights and liberties of the people, the general assembly shall establish and maintain free public schools in this state within ages not excess of twenty-one years as prescribed by law.

There is no further description of what a free public education entails, although one unique feature of the Missouri Constitution is that section 3(b) establishes a minimum *percentage* of public revenues to be dedicated to public elementary and secondary education.

In event the public school fund provided and set apart by law for the support of free public schools, shall be insufficient to sustain free schools at least eight months in every year in each school district of the state, the general assembly may provide for such a deficiency; but in no case shall there be set apart less than twenty five percent of the state revenue, exclusive of interest and sinking fund, to be applied annually to the support of free public schools.

---

<sup>9</sup>There is no scientific evidence for any particular funding weight, a point conceded by one of the plaintiff experts. The decision to weight a poor student at 1.2 or 1.4 times a nonpoor student reflects a normative decision by legislatures as to fairness in resource allocation, rather than an objective assessment of the "cost" of educating the two types of students.



### Percent spent on free public schools Three Approaches, 2004-2006

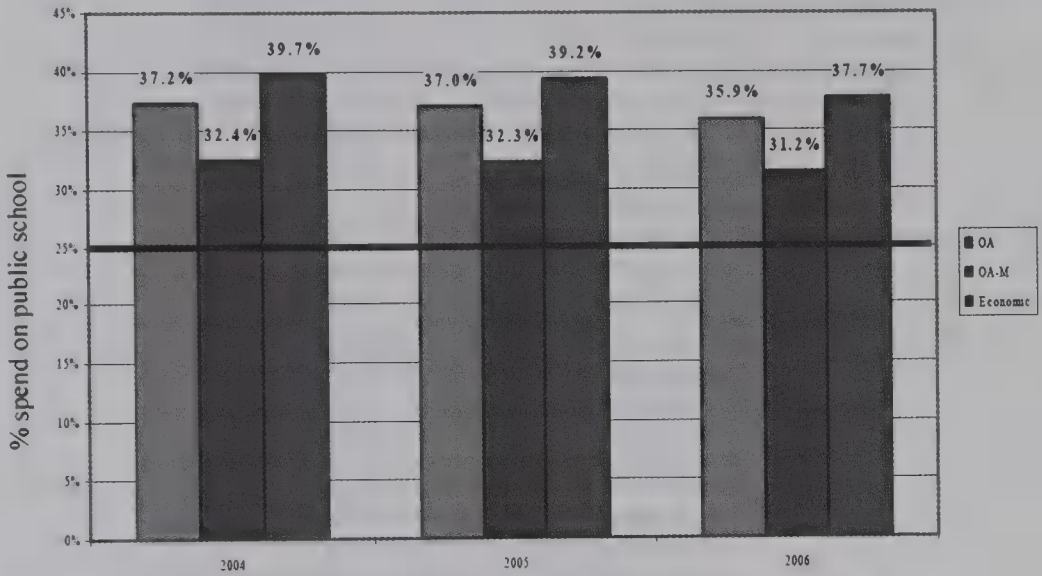


FIGURE 1 Percentage spent on free public schools, three approaches 2004–06. Source: Haslag (2007). OA=Office of Administration, OA-M=Office of Administration-Modified, Economic-Estimated by Haslag.

In his decision, Judge Callahan notes that the requirement is unique to the Missouri Constitution.

This first hurdle is readily met. Figure 1 shows the percentage of state revenues devoted to free public schools under various definitions of education spending and state revenues computed by a University of Missouri Economist, Joseph Haslag, and presented at trial. Under every reasonable measure (and there are alternative ways of measuring both the numerator and denominator), state spending on education far exceeds the 25% threshold.

Whatever K-12 spending may be as a percentage of state revenues, it is useful to know how spending per student figure stacks up against other states and the nation. Figure 2 shows that Missouri ranks somewhat below the middle (32) in current spending per student. In this and some subsequent figures we report not only Missouri but the seven surrounding states. In part this reflects the generally lower wage structure in the Midwest as compared to some other states. Our surrounding states are all in the lower two thirds of the distribution of states, with Missouri in the middle of this group as well. There is no simple way to adjust the data for “cost of living” because a statistically reliable cross-section cost of living index does not exist. However, the National Center for Education Statistics has developed a Current Wage Index (CWI). The CWI is an index that is based on the level of earnings for college-educated workers in the labor market area.

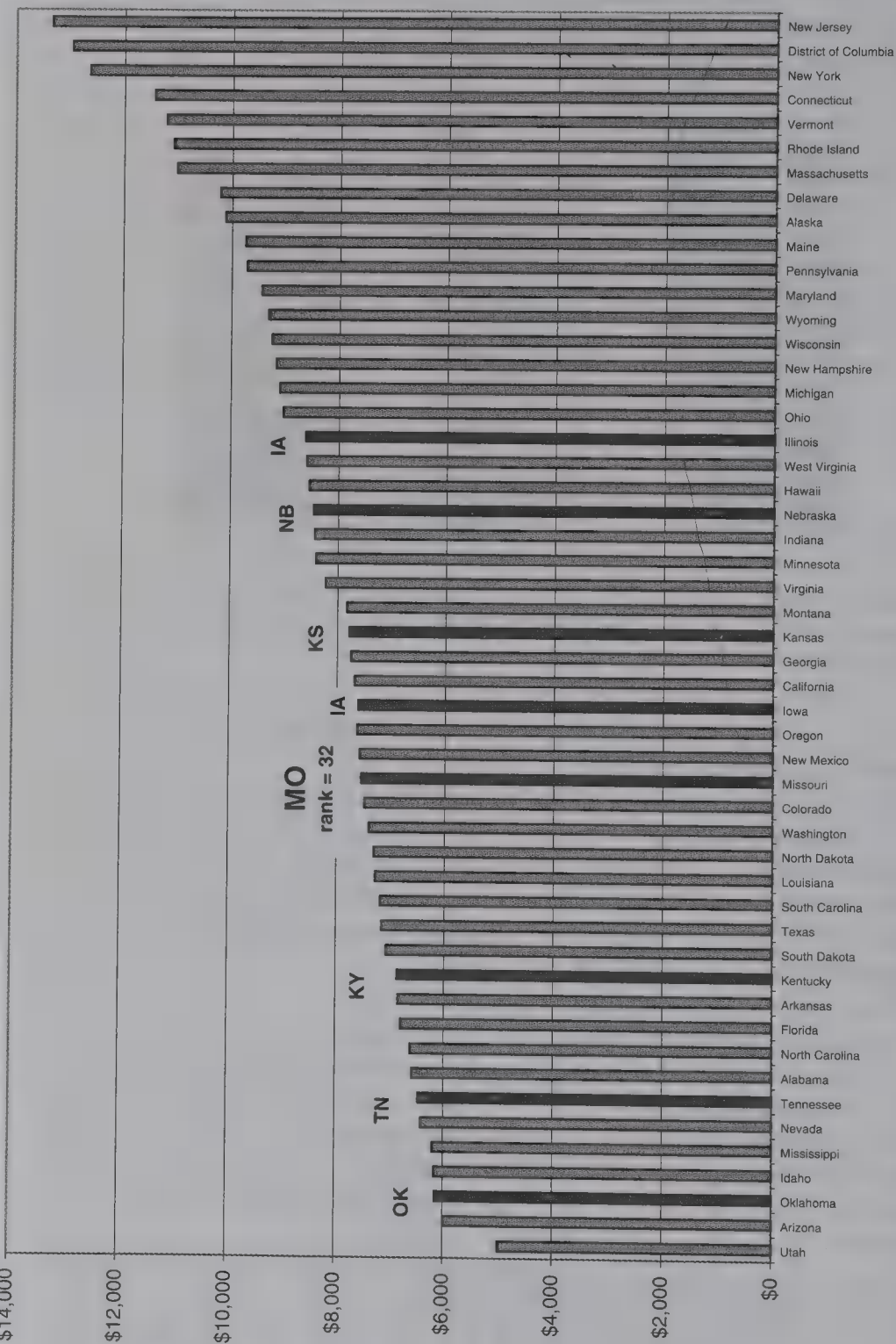


FIGURE 2 Current expenditure per student in Missouri and other states, 2003-04. Source: U.S. Department of Education. Digest of Education Statistics, 2006.

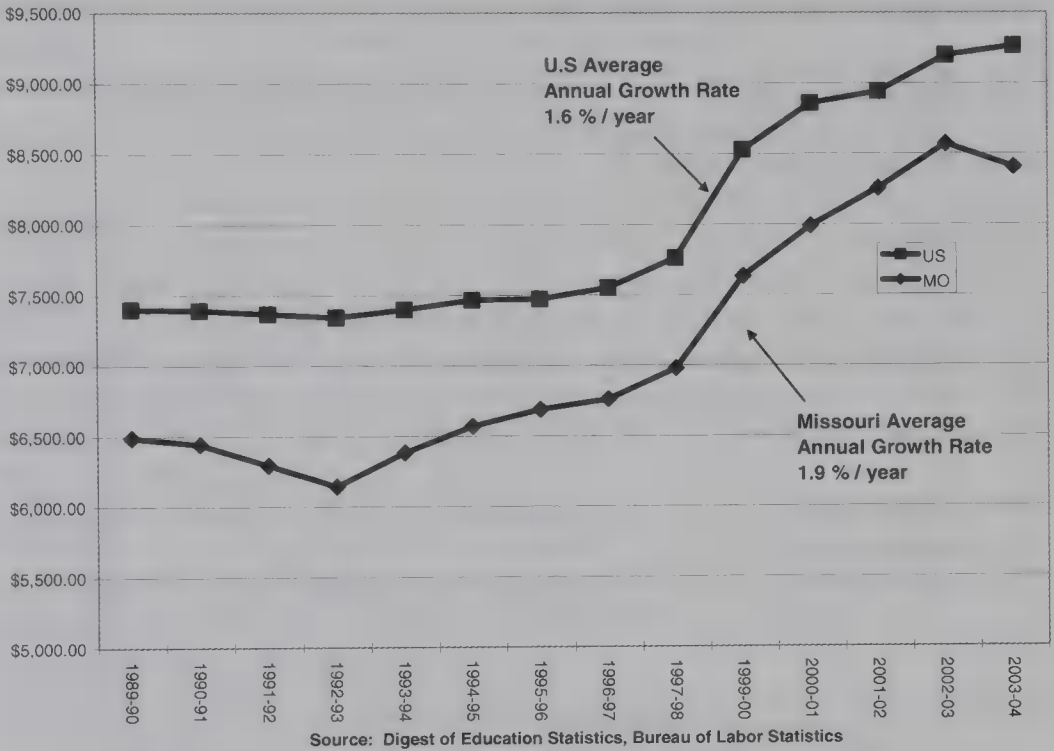


FIGURE 3 Trends in inflation-adjusted current spending per student, 1989–90 to 2003–04.

As such it can be used to compare education labor earnings to those of other college-educated workers. According to the CWI, the general wage structure in Missouri is about 10% below the national average. Because labor costs account for the lion's share of per-student education costs, if we take the CWI as an accurate measure of education labor costs, then a 10% upward adjustment in current spending per student would entirely close the gap between Missouri and the U.S. average.

Although Missouri spending per student is below the national average, over the past decade or more it has been consistently rising in real terms, and at a somewhat higher rate than the national average. Figure 3 reports the average inflation-adjusted spending in Missouri and the United States since the 1989–90 school year. Real per-student spending nationally rose at a 1.6% annual rate, whereas the annualized rate for Missouri was 1.9%.

At issue in this case is how equally these resources are distributed among school districts. Before discussing making any types of interstate comparisons of spending inequality it is important to understand something of the landscape of school districts in Missouri. Relatively speaking, Missouri has a lot of school districts—75 K-8 and 447 K-12, 522 regular school districts in all—many of

TABLE 1  
Enrollment by District Size: 2004–05

<i>Percentile of District Size</i>	<i>% of Students</i>	<i>Cumulative % of Students</i>
Deciles		
10	0.5%	0.5%
20	1.0%	1.5%
30	1.5%	3.0%
40	2.3%	5.3%
50	3.1%	8.4%
60	4.2%	12.6%
70	5.7%	18.3%
80	8.9%	27.2%
90	15.8%	43.0%
100	57.0%	100.0%
Largest 5 districts	16.0%	—
Largest 10 districts	25.8%	—

*Note.* Source: Missouri Department of Elementary and Secondary Education.

which are quite small.<sup>10</sup> In many of our comparisons, we focus on surrounding Midwestern states. Most of these states also have a large number of school districts, many of which are rural. Second, Missouri has a highly skewed distribution of students among these districts: Some have very few students and some have many. Table 1 reports the distribution of students by decile of district size, from lowest to highest. The smallest 10% of Missouri districts enroll just 0.5% of all students. The smallest 20% of districts (i.e., 104 of 524) enroll just 1.5% of public school students. By contrast, the largest 10% enroll over half (57%) of the students. In fact, the largest 10 school districts enroll slightly more than 25% of the students, and the 5 largest enroll 16%.

As discussed previously, most of the early wave of school finance cases focused on how equitably resources were available to school districts. For most of the history of public schooling in America, local school districts primarily were funded by local property taxes. This arrangement produced often dramatic differences among districts in the amount of resources available for educating students. Beginning in California with *Serrano v. Priest* (1971), state courts around the nation decided that such funding disparities were unconstitutional and that the amount of funding available to an individual student should be dependent on the wealth of the state as a whole, not on whether a student was lucky enough to live in a district wealthy in assessable property. In 1993 Missouri’s school finance system

<sup>10</sup>Officially, Missouri has 524 school districts. However, for this study we drop 2—the St. Louis and Pemiscot County Special School Districts—and focus only on regular school districts.



was declared unconstitutional because of such funding disparities among school districts.<sup>11</sup>

These early cases addressed what is commonly termed horizontal equity, that is, the extent to which all students have access to substantially the same level of resources (Berne & Stiefel, 1983). Identical spending per pupil in every district would yield perfect horizontal equity. Such a funding formula, at least tacitly, assumes that every child's education requires identical resources to produce. Most experts in school finance now recognize that some students have characteristics that may require that greater resources be applied to their education. This concept is commonly known as vertical equity, that is, the amount of resources made available is dependent upon an individual student or group of students' identified educational needs. Perfect vertical equity requires that spending be based solely on student need.

Horizontal equity is a straightforward, relatively easily measured concept. Vertical equity may conceptually be straightforward, but in practice it defies precise measurement primarily because the technology of education is not well understood. The state of the art does not allow one to reliably predict the effects of any intervention, input, or combination of inputs.

Complexity is further exacerbated by the imprecise methods of identifying and classifying students with additional needs. Most high-needs classifications tend to be subjective and cover an often broad range of student characteristics. Take, for example, two 12-year-old students from Mexico—one has attended school in rural Mexico for only the equivalent of 4 years, the second transferred from an elite private school in Mexico City. The former barely reads in Spanish; the latter has studied English for 6 years. Both probably would be identified as LEP and in most states would generate identical amounts of extra funding. Should either or both be placed in a bilingual class, language immersion class, English as a Second Language class, or none of the above? There is no definitive science to guide educators in these choices, and each choice carries implications for the required level of spending. Identification and treatment for special education, gifted, economically disadvantaged, and so on, suffer from similar imprecision.

In the quest for vertical equity, state school finance systems sometimes take into account other factors that may affect costs faced by school districts regardless of the characteristics of their students. Small districts may face diseconomies of scale that increase per-pupil costs. Other districts may be located in areas where the costs of goods and services require them to offer higher salaries to attract and retain qualified staff. Others may, because of demographic trends, employ teachers with more-than-average experience who, under current pay schemes, receive higher-than-average salaries.

---

<sup>11</sup> *Committee for Educational Equality v. State of Missouri* (1993).

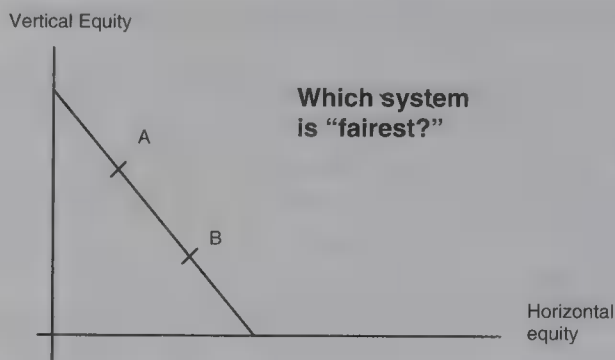


FIGURE 4 The trade-off between horizontal and vertical equity.

Perfect horizontal equity and perfect vertical equity are mutually exclusive. States that choose, at least in part, to condition the level of revenues per pupil a school district receives on the likely needs of its students or other cost factors will fare poorly on measures of horizontal equity. This trade-off is illustrated in Figure 4.

Missouri compensates districts for several cost differences and, as a consequence, tilts in the direction of vertical equity. The current formula adds compensatory funding for districts whose student populations exceed specified thresholds of students eligible for federally subsidized meals (FRL), students identified as limited English speaking (LEP), or handicapped. The previous formula similarly provided extra funding to districts with concentrations of students with greater educational needs. The effect of these policies can be seen in Figures 5 through 7.

Figure 5 plots weighted measures of school spending inequality for Missouri and seven surrounding states. Several patterns emerge. First, inequality for Missouri is considerably higher than all of the surrounding states. Second, inequality dropped sharply in the wake of the 1993 court decision, but the decline stopped by the late 1990s. Thus, by a measure of horizontal equity, Missouri compares poorly with surrounding states.

What about vertical equity? Here the story changes. Figure 6 shows the correlation between per-pupil spending and student poverty in Missouri and surrounding states for 1990 to 2002, the latest period for which there is available data. Figure 7 displays the correlation between district per-pupil spending and the percentage of minority students in the district in Missouri and surrounding states. These graphs clearly demonstrate that Missouri spends significantly more on students who are more likely to have greater-than-average needs. This is a consequence of deliberate policy choices to attempt to compensate districts that likely face higher-than-average costs because of the characteristics of students attending their schools. Most states and the federal government provide some form of compensatory funding, but as can be seen from Figures 6 and 7, Missouri

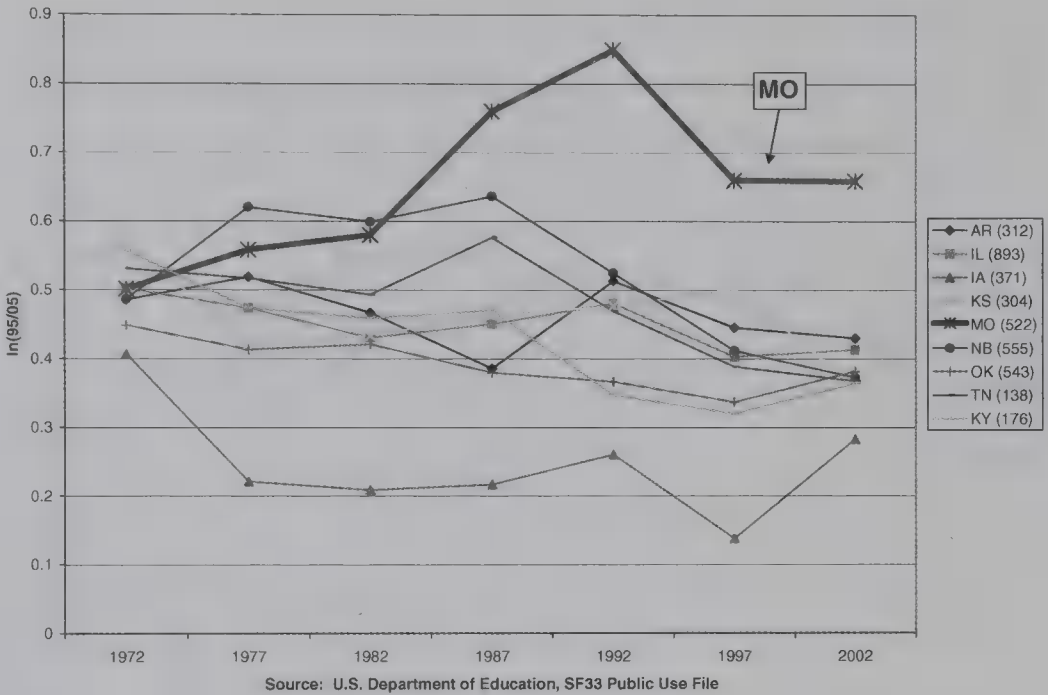
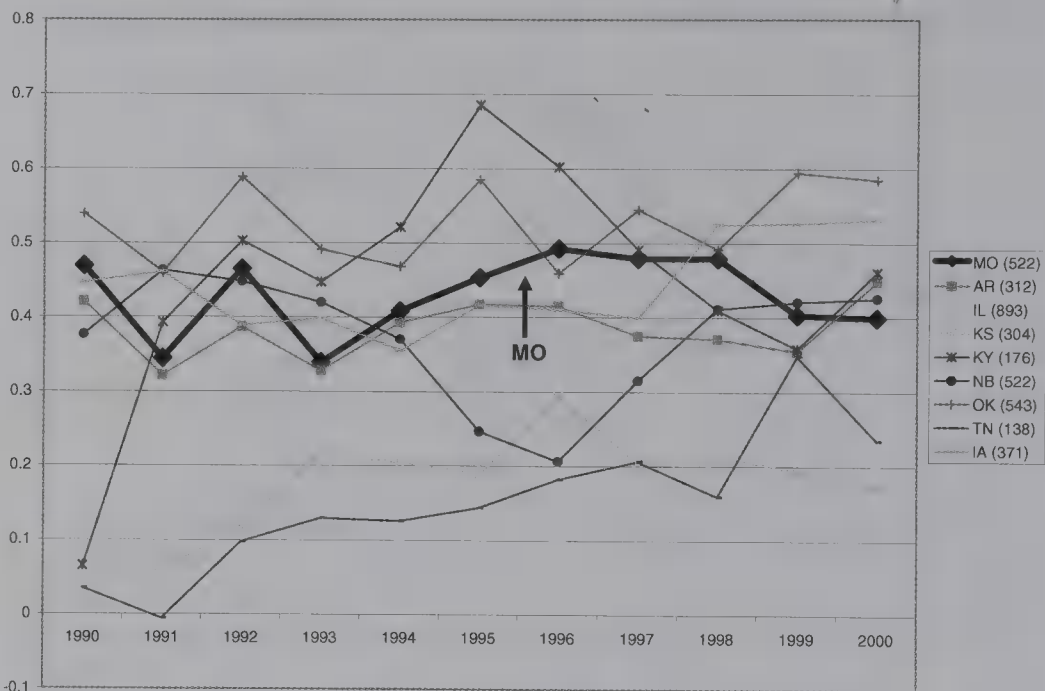


FIGURE 5 Measured inequality in current spending per student in Missouri and surrounding states: 1972–2002. *Note.* The number of school districts (2002) is in parentheses in the legend. Source: U.S. Census Bureau, Elementary and Secondary School System Finance Data Files (F-33). Inequality measure ( $\ln[95\text{th}/5\text{th percentiles}]$ ).

has one of the more aggressive policy structures in this regard in the region. To most observers this would be a good thing, but Plaintiffs expert witnesses Richard Salmon and Lisa Driscoll concluded that Missouri school finance was inequitable and generally getting worse (Salmon & Driscoll, 2006).

### *“Adequacy” in Relation to Student Test Scores*

A critical claim of plaintiffs in the Missouri and other adequacy cases is that there is a constitutional standard of adequate spending defined in relation to a certain set of student skills. In other words, to make sure that students have a certain set of academic skills, it is necessary to spend at least  $X$  dollars per student (where  $X$  can be adjusted based on student need). This, in turn, assumes that there is a statistically reliable and causal positive relationship between school spending and student achievement. Let us consider each of these in turn. By “statistically reliable,” we mean that the relationship is stable because the student achievement associated with a given level of spending is highly predictable. To put it simply, an expenditure level of \$8,000 per student would be consistently associated with



**FIGURE 6** Correlation between current spending per-pupil spending and student poverty in Missouri and surrounding states: 1990–2000. *Note.* The number of school districts (2002) is in parentheses in the legend. Source: National Center for Education Statistics, Longitudinal School District Fiscal-Nonfiscal Data File. Inequality measure (ln[95th/5th percentiles]).

a given level of student achievement, and an expenditure level of \$10,000 per student would be associated with a higher, predictable level of achievement. In the world of No Child Left Behind, this reliability is taken to mean “if I spend  $X$  per student, I can expect to see a proficiency rate of  $A$ , and if I spend  $1.25 X$ , I can expect to see a proficiency rate of  $B$ , where  $B$  is bigger than  $A$ .”

The second condition is equally important. Even if we found a positive and stable statistical relationship between district spending and student achievement, it does not mean that the former caused the latter. For example, it may be that high-spending districts are also districts with more affluent, well-educated parents. It is well established in the research literature that the most powerful predictors of student achievement are family background factors, particularly parents’ education (Hoxby, 2001). Although school-age children spend roughly 1,100 hr per year at school, they spend thousands of hours more at home. Moreover, parental nurturing during the preschool years also plays an important role in children’s development (Armor, 2003). Many studies have demonstrated that the “summer melt” is much larger for children from low- as compared to high-income families (Cooper, Nye, Carlton, Lindsay, & Greathouse, 1996). The bottom line is that,



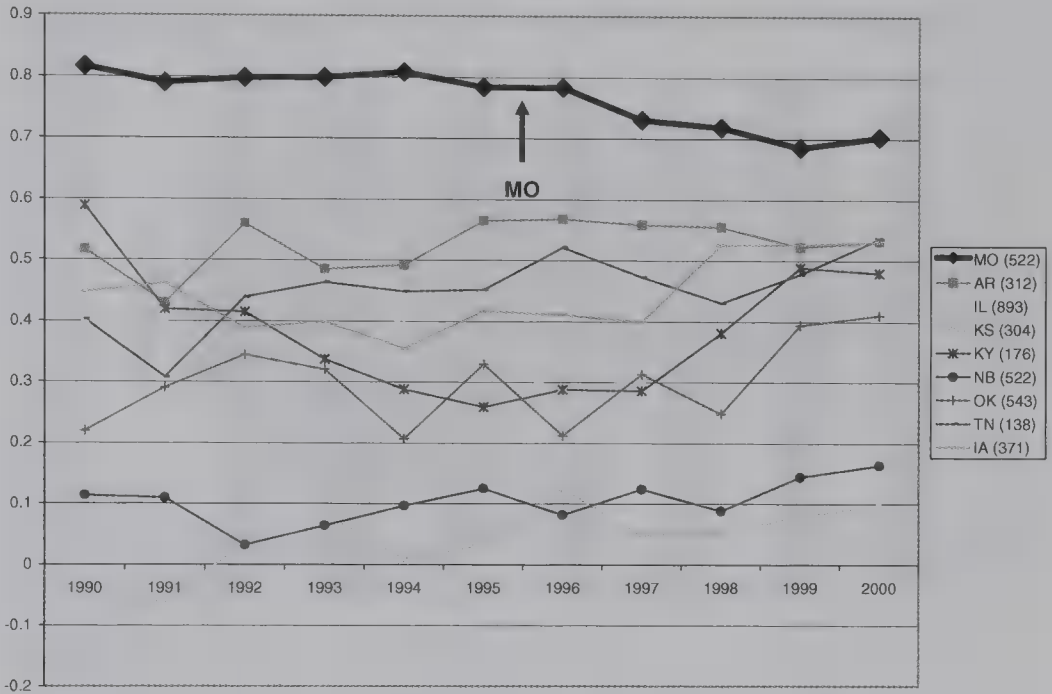


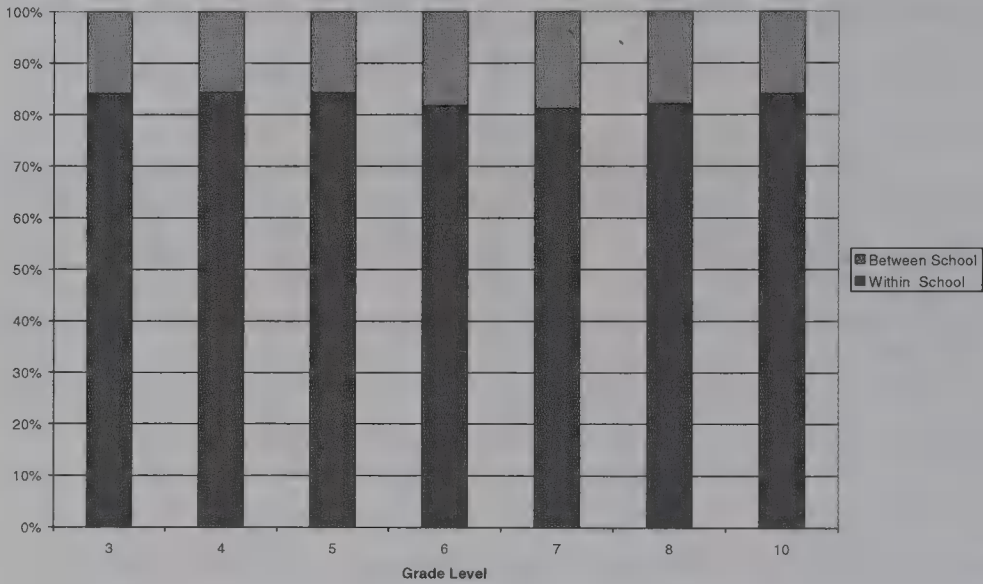
FIGURE 7 Correlation between current spending per-pupil spending and percentage minority in Missouri and surrounding states: 1990–2000. *Note.* The number of regular school districts is in parentheses in the legend.

on average, higher income families make larger human capital investments in their children at home. If high-income families cluster in high-spending school districts, this will produce a positive relationship between spending per student and student achievement even if school spending has no causal effect on student achievement.

In the court case, one of the authors made an exhaustive examination of student-level data on the statewide Missouri Assessment Program, the state assessment used in public schools. We do not report all of those results but focus on just a few charts.<sup>12</sup> The “remedies” in school finance trials focus on changes in how the states fund school districts, and “equity” and “adequacy” are defined in terms of school districts, not students. The presumption is that interdistrict gaps in student achievement are a major source of student achievement gaps. In fact, the data in Figure 8 show that interdistrict gaps in student achievement are a minor source of achievement inequality for students and the vast majority of inequality in achievement is within school districts. Here we report what are called analysis

<sup>12</sup>The complete report is available at <http://www.schoolchoiceformissouri.org/trial/trialselecteddefense.html>

A. Math



B. Communication Arts

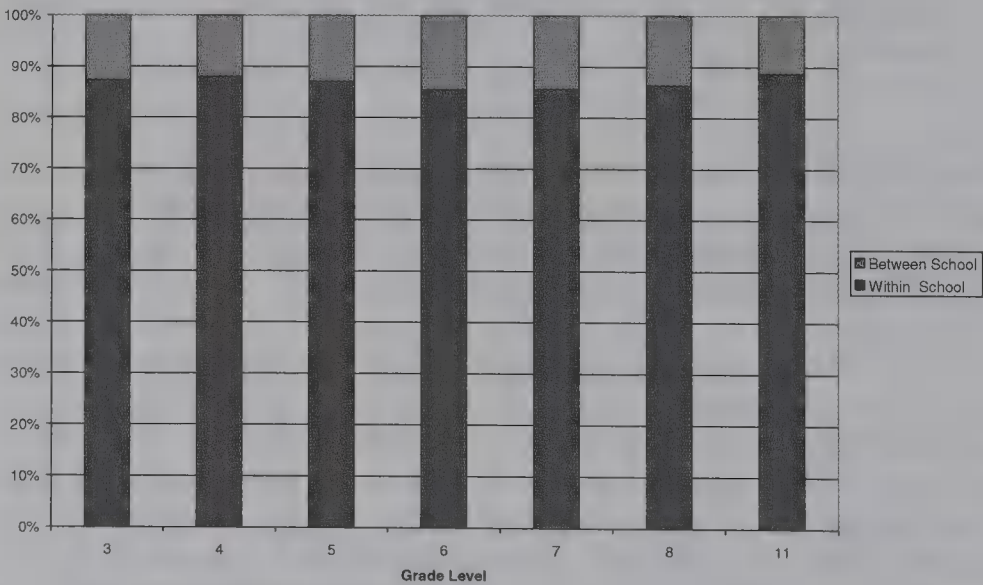


FIGURE 8 Sources of variation of 2006 student MAP test scores: Within and between districts by grade level.

of variance (ANOVA) decompositions. Basically each bar represents total student achievement inequality (100%) for each test (communications arts, math) at each grade level (3–8, high school). Total inequality of achievement at any grade can be broken into the sum of two components: inequality between districts and inequality

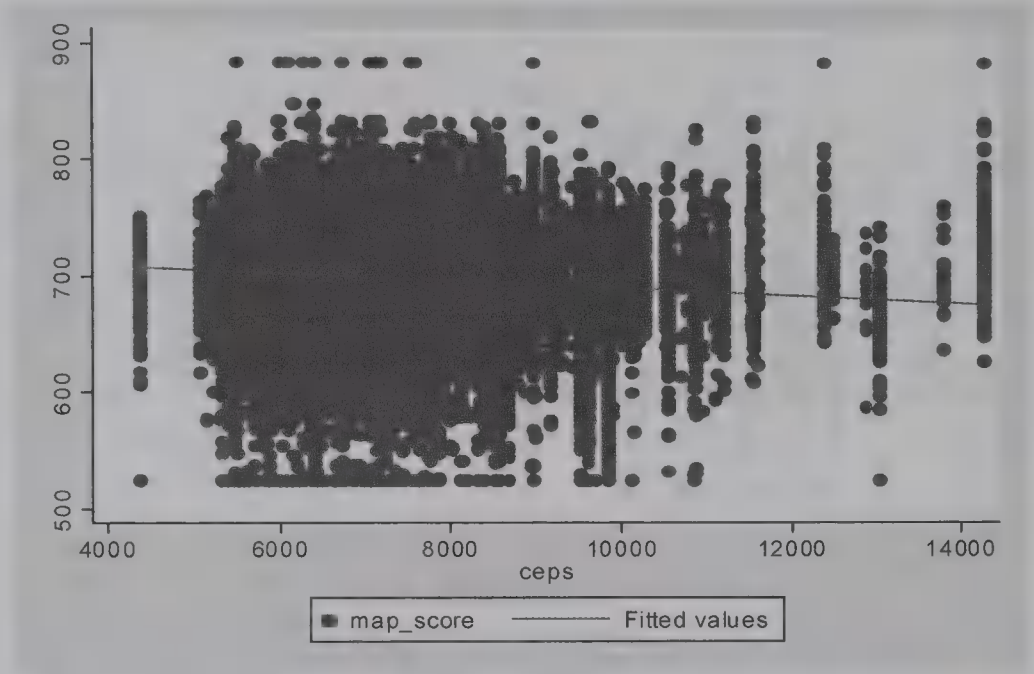


FIGURE 9 2006 Grade 8 Math MAP test scores and current expenditures per student. *Note.* Source: Missouri Department of Elementary and Secondary Education. All reported District MAP scores with at least 25 valid test scores in beginning and ending year.

within the districts.<sup>13</sup> The share of variation between districts is the percentage of inequality that would disappear if all school districts had the same average test score. As both charts show, the vast majority of inequality is *within* rather than *between* districts. At most only 15 to 20% of math inequality and 10 to 15% of communication arts inequality is between districts. The overwhelming share of inequality is within districts. Thus, even if equalizing spending across districts eliminated all interdistrict inequality of spending, the vast majority of inequality would remain.<sup>14</sup> In spite of the fact that the vast majority of inequality is within rather than between districts, it is typical in school finance trials for experts on both sides to focus on average district achievement and average district spending.

In light of the minor role played by interdistrict variation in inequality, we believe the proper area of focus is student level achievement data. We have analyzed such data at all grade levels. The interested reader is referred to the full trial report. However, Figures 9 and 10 depict a consistent pattern. Each dot in these

<sup>13</sup>Stated more formally, ANOVA decomposes total variation in student test scores into two components: variation in mean achievement between districts, and variation within the district around the district mean.

<sup>14</sup>A similar decomposition can be made within and between schools. At least 80% of achievement inequality is within schools.

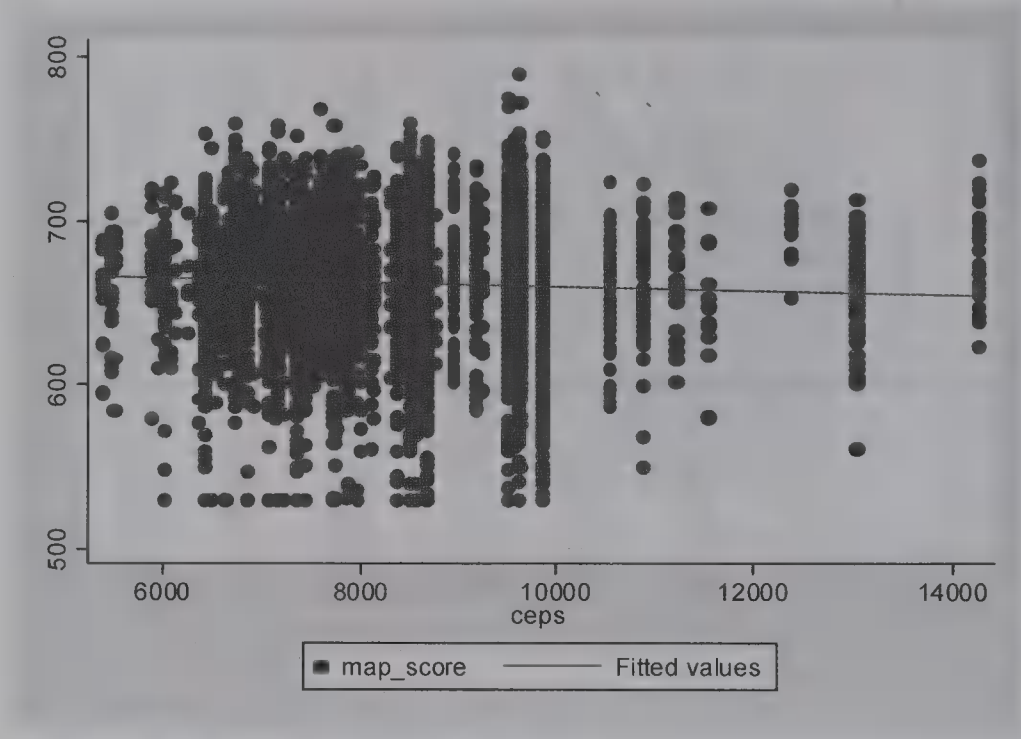


FIGURE 10 2006 Grade 8 Math MAP test scores and current expenditures per student: African American students who are FRL eligible.

diagrams depicts a student (or more commonly, a “stack” of students with identical scores). The first diagram shows eighth-grade math scores for all students in the state (roughly 71,000). Along the horizontal axis is displayed the average current spending in the district.<sup>15</sup> If district spending per student had a powerful effect on achievement, one would expect to see a positive relationship between achievement and spending. Instead what one finds is a “cloud” of scores with no apparent relationship. In fact, a regression line fit through these data has a slightly negative slope. Moreover, these data reinforce the finding just discussed, namely, that the vast majority of inequality is within rather than between districts. Because every student in a district is assigned the same value of spending per student, these score distributions are stacked district by district, ranked by spending. Not only is there no positive relationship between spending and average achievement, but

<sup>15</sup>One important limitation of school finance data in Missouri and most other states is that spending per student can only be measured at the district level. Thus all students in the same district are assigned the same value of per-pupil spending rather than a measure of the actual resources expended on them. Although spending per student is generally not available at the school building level, some researchers have been able to secure school-level resource data in some districts (Iatarola & Stiefel, 2003; Roza, Guin, Grosse, & Deburgomaster, 2007; Roza & Hill, 2004). They find considerable intradistrict inequality, arising primarily from differences in average teacher pay between schools.



there is no tendency toward compression either. High-spending districts are just as unequal as low-spending districts.

It might be argued that a chart such as this, which plots all students, confounds the effect of student socioeconomic status (SES) and race. It is well known that low-SES students on average have lower test scores. The same is true for African American students. Thus, if districts with above-average spending per student also have above-average shares of poor or African American students then a positive effect of spending on student achievement would be confounded by this SES or racial effect. To neutralize the latter, we plot in Figure 10 identical data just for African American students who are free and reduced-price lunch eligible, a standard measure of student poverty (roughly 8,400 such students were tested in eighth grade). Once again, there is no evidence of a positive effect of district spending.

## CONCLUSION

In this article, we have reviewed some of the evidence presented in the recent school finance trial Committee for Educational Equality (CEE) II. Although many states have experienced such litigation, the Missouri experience was unusual in two respects. First, three taxpayers were allowed to intervene on behalf of the defense. This opened the door to a more vigorous rebuttal by the defense experts and, more important, permitted the defendants to raise important questions concerning the efficiency of school spending and school reform. It is a model that should be considered in other states. Second, the trial at the circuit court was a complete victory for the defense. Although recent school finance cases have not gone well for the plaintiffs, this one was a particularly sharp loss. We believe the latter followed as a direct consequence of the former—the taxpayer intervention played a strong role in sharpening the defense.

We reviewed some of the key evidence presented to the court concerning equity in the distribution of educational resources and the feasibility of establishing a level of “adequate” resources with reference to student achievement. Like most states, in recent decades Missouri has consistently increased the real spending per student in K-12 education. In fact, real per-student spending has risen faster than the national average. When we consider how equitably those resources were distributed, we showed that there are potentially important differences in how one measures inequality. When considered simply in terms of the inequality of spending per student (horizontal equity), Missouri does not compare favorably to surrounding states. However, with vertical equity measures that adjust spending for student need, Missouri compares very favorably. Finally, we show that efforts to specify an “adequate” level of K-12 spending per student by reference to student test scores is a hopeless endeavor. It is simply not possible to identify a statistically reliable relationship between district spending and student achievement.

In closing, we believe that the data show that school finance litigation is much too blunt an instrument to address issues of student achievement gaps. There is simply no evidence that court-induced changes in school finance play an important role in changing student achievement gaps. Unfortunately, the remedies being suggested to the courts—changing formulas for state aid to school districts—have virtually no relationship to student achievement gaps because the vast majority of student achievement inequality is within rather than between districts. If school finance systems are to be challenged in courts, we believe that the student should be the focus of judicial remedies rather than “school districts.” True equity and efficiency is more likely to be achieved when state dollars are attached to students, whose resources travel with them as their parents choose the best school to fit their needs.

## REFERENCES

- Armor, D. J. (2003). *Maximizing intelligence*. Edison, NJ: Transaction.
- Atkins, C. (2007). *Appropriation by litigation: Estimating the cost of judicial mandates for state and local education spending*. Washington, DC: Tax Foundation.
- Augenblick, J., & Myers, J. (2003). *Calculation of the cost of an adequate education in missouri using the professional judgment and the successful school district approaches*. Missouri Coalition for Adequacy. Retrieved February 2005 from <http://www.msbanet.org/pdf/adequacy.pdf>
- Berne, R., & Stiefel, L. (1983). *The measurement of equity in school finance: Conceptual, methodological, and empirical dimensions*. Baltimore: Johns Hopkins University Press.
- Berry, C. (2007). The impact of school finance judgments on state fiscal policy. In M. West & P. Peterson (Eds.), *School money trials: The legal pursuit of educational adequacy* (pp. 213–242). Washington, DC: Brookings Institution Press.
- Campbell County School District, State of Wyoming, et al. v. State of Wyoming: Diana J. Ohman, State Superintendent, et al. No. 94–136, No. 94–137, No. 94–138, No. 94–139, No. 907 p. 2d 1238; 1995 Wyo. LEXIS 203.
- Clark, M.A. (2003, October). *Education reform, and student achievement: evidence from the kentucky education reform act*. Princeton, NJ: Mathematica Policy Research.
- Committee for Educational Equality v. State of Missouri*, No. CV190-1371CC, slip op. (1993).
- Cooper, H., Nye, B., Carlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Education Research*, 66(3), 227–268.
- Franck, M. (2007, October 10). Taxpayers pay tab in school-funding lawsuit. *St. Louis Post-Dispatch* (p. B1).
- Guthrie, J. W., & Springer, M. G. (2007). Alchemy: Adequacy advocates turn guesstimates into gold. *Education Next*, 7(1), 20–27.
- Guthrie, J. W., Springer, M. G., Rolle, A. R., & Houck, E. A. (2007). *Modern education finance and policy*. Boston: Allyn & Bacon.
- Hanushek, E. A., & Rivkin, S. G. (1997). Understanding the 20th century growth in U.S. school spending. *Journal of Human Resources*, 32(1), 35–68.
- Haslag, J. (2007). *Spending on free public schools: Expert report* Retrieved from <http://www.schoolchoiceformissouri.org/trial/pdf/Haslag%201011%209-18-07%20revised.pdf>
- Hoxby, C. (2001). If families matter most, where do schools come in? In T. M. Moe (Ed.), *A primer on America's schools* (pp. 89–125). Palo Alto, CA: Stanford, Hoover Institute Press. Iatarola, P., &

- Stiefel, L. (2003). Intradistrict equity of public education resources and performance. *Economics of Education Review*, 22(1), 69–78.
- Murray, S. E., Evans, W. N., & Schwab, R. M. (1998). Education-finance reform and the distribution of education resources. *The American Economic Review*, 88(4), 789–812.
- Podgursky, M., & Springer, M. (2006). K-12 public school finance in Missouri: An overview. *Regional Economic Development*, 2(1), 3–50.
- Roza, M., Guin, K., Grosse, B., & Deburgomaster, S. (2007, Fall). Do districts fund schools fairly? *Education Next*, 7(1), 68–74.
- Roza, M., & Hill, P. (2004). How within district spending inequities help some schools to fail. *Brookings Papers on Education Policy* (pp. 201–227). Washington, DC: The Brookings Institution.
- Rose v. Council for Better Education* (KY). 790 SW2d 186, 60 Ed. Law Rep. 1289 (1989).
- Salmon, R. G., & Driscoll, L. G. (2006, December). *An analysis of fiscal equity provided by the Missouri System for Financing Public Schools: 1992-93 to 2004-05, Expert report*. Retrieved from <http://www.schoolchoiceformissouri.org/trial/pdf/Salmon%20Driscoll%20MO83285.pdf>
- San Antonio Independent School District et al. v. Rodriguez et al., Appeal from the United States District Court for the Western District of Texas No. 71–1332. Argued October 12, 1972, Decided March 21, 1973.
- Serrano v. Priest*, 5 Cal. 3d. 584, 96 Cal Rptr. 601, 487 P. 2d. 1241 (1971).
- Springer, M. G., Liu, K., & Guthrie, J. W. (2007). *The impact of education finance litigation reform on resource distribution: Is there anything special about adequacy?* (Policy Working Paper). Nashville, TN: Peabody Center for Education.

# What Do Cost Functions Tell Us About the Cost of an Adequate Education?

Robert Costrell  
*University of Arkansas*

Eric Hanushek and Susanna Loeb  
*Stanford University*

Econometric cost functions have begun to appear in education adequacy cases with greater frequency. Cost functions are superficially attractive because they give the impression of objectivity, holding out the promise of scientifically estimating the cost of achieving specified levels of performance from actual data on spending. By contrast, the opinions of educators form the basis of the most common approach to estimating the cost of adequacy, the professional judgment method. The problem is that education cost functions do not in fact tell us the cost of achieving any specified level of performance. Instead, they provide estimates of average spending for districts of given characteristics and current performance. It is a huge and unwarranted stretch to go from this interpretation of regression results to the claim that they provide estimates of the minimum cost of achieving current performance levels, and it is even more problematic to extrapolate the cost of achieving at higher levels. In this article we review the cost-function technique and provide evidence that draws into question the usefulness of the cost-function approach for estimating the cost of an adequate education.

Econometric cost functions have begun to appear in education adequacy cases with greater frequency. Although previously considered too technical for courts to understand, recent litigation in Missouri featured separate cost-function estimates commissioned by each of two plaintiffs. A prior Texas court case presented results

---

This article was a conference paper for *From Equity to Adequacy to Choice: Perspectives on School Finance and School Finance Litigation*, Show-Me Institute and the Truman School of Public Affairs of the University of Missouri, October 30, 2007.

We acknowledge the support of the Missouri Show-Me Institute. The usual disclaimers apply.

Correspondence should be sent to Robert M. Costrell, Department of Education Reform, 201 Graduate Education Building, University of Arkansas, Fayetteville, AR 72701. E-mail: costrell@uark.edu



from two dueling cost studies commissioned by the opposing sides (Gronberg, Jansen, Taylor, & Booker, 2004; Imazeki & Reschovsky, 2004b). This increased use of the cost-function methodology likely reflects growing skepticism about other methods typically used to estimate the cost of providing an adequate education. In particular, the professional judgment (PJ) method has begun to lose favor.<sup>1</sup> In this approach, panels of educators design prototype schools that they believe will provide adequate educational opportunities, and then the consultants hired to conduct the study attach costs to these prototypes. Even a sympathetic trial judge in Massachusetts concluded that the PJ study submitted there was “something of a wish list” (Costrell, 2007, p. 291). Hence, although PJ studies are invariably included, recent finance cases have attempted to bolster these with econometric cost functions.

Cost functions are superficially attractive because they appear objective, holding out the promise of scientifically estimating the cost of achieving specified levels of performance from actual data on spending instead of relying on opinions, as do PJ estimates. In keeping with this perception, a group of education finance specialists began arguing that econometric cost functions are the most scientifically valid method to determine the cost of adequacy. To make this argument, they asserted that the methods for estimating cost functions in the private sector—where competition tends to drive out inefficient producers—could be readily adapted to public education. They prepared estimates for legislative committees and courts in states such as New York, Texas, Kansas, and Missouri and published their work in academic journals. The problem, we argue, is that education cost functions do not in fact tell us the cost of achieving any specified level of performance, as claimed.

This is not to say that cost functions tell us nothing. They do provide estimates of average spending for districts of given characteristics and indicate how spending varies by these characteristics in the specific state. For example, they may tell us that in state X, per-pupil spending averages *Y* thousand dollars for districts with a certain percentage of free or reduced-price lunch eligible (FRL) students or of Black students and that the average rises or falls by *Z* dollars as these percentages change. Regression equations provide a useful summary of such patterns. By extension, including measures of performance (e.g., test scores) as a variable permits summarizing what the average spending is for districts with given demographics and performance levels.

However, it is a huge and unwarranted stretch to go from this modest interpretation of regression results to the far more extravagant claim that these provide estimates of the cost of achieving any given performance level for districts of

---

<sup>1</sup>The alternative methods are discussed in Ladd, Chalk, and Hansen (1999). Particular attention to the use of cost functions can be found in Gronberg et al. (2004), Duncombe (2006), and Baker (2006a). Critiques can be found in Hanushek (2006, 2007).

given demographics. There are two key heroic assumptions that are required: The estimates of *average spending* among comparable districts can be adjusted so that they reflect the *minimum efficient cost*<sup>2</sup> to generate current performance levels, and the estimated variation in average spending across districts with different performance levels can be used to extrapolate the costs of raising performance to levels not currently observed by comparable districts.

As we show in this article, the method typically used to convert average spending figures into estimates of efficient cost accomplishes nothing of the sort. For that reason, there is no foundation for interpreting spending variations across districts with different demographics as the required spending premiums for demographic groups. Finally, the estimated relationship between “cost” and performance is highly unreliable; it is typically estimated with huge imprecision, wide sensitivity to model specification, and by methods that often fail to eliminate statistical bias. As a result, the cost estimates for raising performance to target levels have no scientific basis.

None of this should be surprising. The recent push for experiments in education research is just one of many indications of the difficulty of estimating the effects of resources on student learning. Why would we need experiments if we could just use average district spending and average student test scores, as do cost functions, to estimate the effect of resources on achievement? Decades of research have repeatedly failed to find a systematic empirical relationship between average spending and performance. It would be quite noteworthy if a handful of recent spending equations were to suddenly have found a relationship that had eluded decades of previous investigation. This simply is not the case. The deeper reasons for this and the consequences thereof are the subject of this article.

## THE BASIC PROBLEM: THE CLOUD

The logic behind regression-based estimates of the cost of adequacy is seemingly compelling. Why shouldn't we be able to use data on district spending and student test performance to estimate the costs of achieving a given outcome goal?

The dimensions of the difficulty with this are easiest to see by looking at the simple relationship between spending and performance. Figure 1 shows a plot of spending and performance in 2006 for the 522 districts of Missouri. The vast majority of districts lie in a solid cloud of spending between \$5,000 and \$8,000 per student and with average achievement on the Missouri Assessment Program (MAP) tests between roughly 700 and 800. At virtually any spending level in the

---

<sup>2</sup>To an economist, this is a doubly redundant phrase, as “cost” implies efficiency, which in turn implies minimum spending necessary to achieve a given outcome. Because this usage may not be universal, we use this phrase for clarity and emphasis.

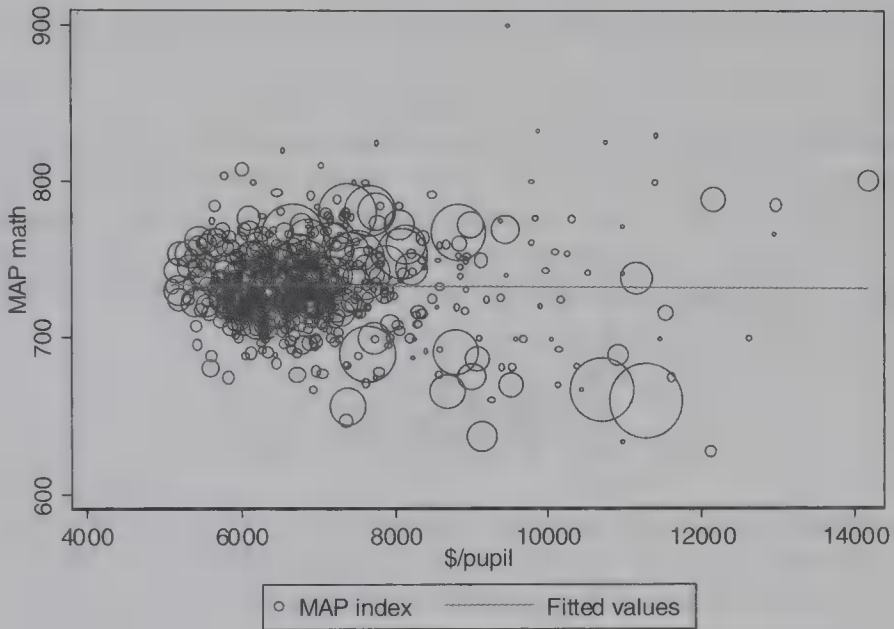


FIGURE 1 Missouri district average eighth-grade mathematics scores and district spending: 2006.  
MAP = Missouri Assessment Program.

range of \$6,000 to \$8,000 there are some districts below 700 points and some at more than 800. This blob of data illustrates the two dimensions of the difficulty previously referred to: average spending differs greatly from minimum spending at any given performance level, and there is no apparent association between average performance and average spending in this group.

There is a smaller number of districts spending more than \$9,000 but still no obvious pattern of being high or low on the math tests. In addition, the size of the circles indicates the student populations. Some large districts are above average in performance, whereas others are below average. The two large and high-spending districts that stand out are Kansas City and St. Louis. Both are noticeably below average in student performance.

Taking all the districts together, the line in the picture shows that the simple relationship between spending and achievement is essentially flat. Even if, on average, there is a small relationship between average spending and average achievement, either positive or negative, the relationship is very weak. That is the fundamental challenge. How can one project the spending necessary to improve student performance to any level when the available data show little tendency toward higher achievement when given extra funds?

Districts of course differ in a variety of dimensions other than spending, leading to a considerable amount of analytical effort to control for other factors in order to

uncover any systematic influence of spending. The basic question is whether other factors that might affect performance, such as poverty levels, can be used to sort districts out of the cloud of Figure 1 such that a pattern with spending emerges. Extensive efforts to do this, beginning with the Coleman Report (Coleman et al., 1966), have been quite unsuccessful. These efforts, generally labeled estimation of production functions, have concentrated specifically on different backgrounds of students and have attempted to standardize for family inputs that are outside the control of schools (Hanushek, 2003).

### THE COST FUNCTION APPROACH

The estimation of cost functions approaches this problem in a slightly different manner than most research exploring the relationship between spending and achievement. It focuses on how achievement levels determine spending, as opposed to how spending determines achievement. When put in terms of the determinants of spending, other things logically enter the analysis. First, districts might differ meaningfully in the prices that they face for inputs, particularly teachers. The price for teachers and other college graduates can be quite different for one district than for another because of the labor markets in which they compete. If districts must pay higher prices to obtain the same quality of resources, then omitting price differences could bias the estimated relationship between achievement and spending. Second, cost functions, similar to production functions, must account for possible variation in resource needs arising from students who have fewer resources at home and thus may require more resources at school, on average, to achieve the same level of performance. Again, if need differences are omitted from cost functions, the estimated relationship between achievement and spending may be biased. Third, districts may differ in the efficiency with which they use their funds. Two districts with similar spending, similar prices, and similar needs might achieve quite different outcomes, based on the efficiency with which they use their dollars to produce the outcome in question. To isolate cost, these estimates must address differences in efficiency.

The underlying premise of the cost-function estimation is that correcting for price differences, the demands of different student bodies, and the efficiency of district spending will yield a clear relationship between achievement and the spending that is required to achieve each level of performance. This relationship then permits identifying the spending required to achieve any chosen level of student achievement.

Do these corrections work?

To answer this question, we trace through some specific cost-function analyses. We focus on those submitted in the Missouri court case because the data were



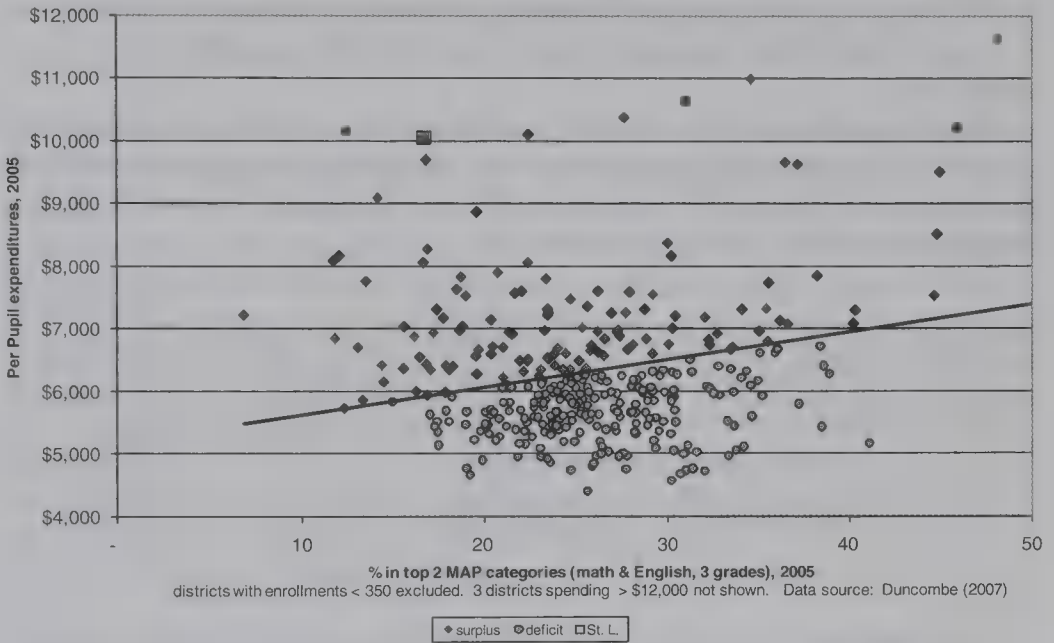


FIGURE 2 Expenditures versus performance in Missouri, 2005.  
MAP = Missouri Assessment Program.

readily available for purposes of replication and analysis.<sup>3</sup> However, the issues identified here apply to the entire genre of cost functions based on the “efficiency control” approach.<sup>4</sup>

Figure 2 presents data on spending and performance, similar to Figure 1.<sup>5</sup> The performance measure, for 2005, is a composite of each district’s performance on the state assessments—specifically the percentage of students in the top two categories (out of five) on the math and communications arts exams across three grades. Unlike Figure 1, Figure 2 places spending, to be determined by achievement and school factors, on the vertical axis. The figure again

<sup>3</sup>See Baker (2006b) and Duncombe (2007). Baker was retained by the main group of plaintiff districts, the Committee for Educational Equality, and Duncombe was retained separately by the City of St. Louis. For the defense, Costrell was retained by the Attorney General of Missouri and Hanushek by the Defendant Intervenors (Shock, Sinquefield, and Smith).

<sup>4</sup>In addition to some of the studies previously cited, a partial list would also include Duncombe and Yinger (1997, 2000, 2005, 2007), Imazeki (2008), Imazeki and Reschovsky (2004a, 2004b), and Reschovsky and Imazeki (2003). Imazeki and Reschovsky, in their various publications about costs in Texas, alternately used an efficiency index derived from a data envelope analysis, including a Herfindahl index, or ignore the issue.

<sup>5</sup>Figures 2 to 5 are based on the Duncombe data and analyses. The Baker data and analyses are very similar, and the corresponding figures are available upon request. Both studies pool data across several years, although these diagrams depict only 1 year.

shows there is a wide range of spending observed at any given level of performance.<sup>6</sup> As a result, the line fitted through these data exhibits a very weak relationship.<sup>7</sup>

What a cost function tries to do is to go beyond this simple (weak) relationship to estimate for a district of *given characteristics* the *minimum expenditure* required to meet some *target performance level*. This can be logically broken down into three steps in constructing the cost estimates:

1. *Control for district characteristics, so that "likes" can be compared with "likes"*: As mentioned, one reason for the wide range of spending is that districts differ in characteristics, such as demography, school size, input prices, and variables thought to affect efficiency. The variation in spending among districts with comparable scores is partially related to these differences. Cost functions statistically control for demographics and other district characteristics with the conventional technique of multiple regression, discussed in the next section.
2. *Purge inefficiency from the estimates of spending*: This is the key step in converting a spending function to a "cost" function. It does so by standardizing the values of the "efficiency controls" used in the first step. If successful, this procedure would identify the minimum expenditure required to perform at the *current level*.
3. *Estimate the cost of raising performance to the target level*: This involves using the estimated relationship between cost and performance to predict the cost associated with increasing performance to a set goal. It requires a reliable estimate of the relationship between cost and performance from the first step.

As we show subsequently, the cost-function methodology does not succeed in this agenda. To understand the issues more fully, we provide a detailed discussion of these steps: (1) controlling for district characteristics, (2) purging inefficiency from average spending, and (3) estimating the additional cost for additional performance.

---

<sup>6</sup>Similarly, there is a very wide horizontal range: At any given spending level, performance varies widely.

<sup>7</sup>In fact, if these data suggest any relationship at all, it is U-shaped, rather than linear, which means a negative relationship between performance and spending over the lower ranges of performance and a positive relationship over the higher ranges. The linear relationship has an  $R^2$  of only 4%, the portion of the variation in spending accounted for by variations in performance. A quadratic relationship, depicting the U-shaped curve, provides a much better fit, with an  $R^2$  of 30%.

## THE ECONOMETRICS OF SPENDING EQUATIONS: CONTROLLING FOR DISTRICT CHARACTERISTICS

The basic technique of linear regression is illustrated by the line through the data in Figure 2. Each point on the line represents the best estimate for *average* spending among districts with any given test score. Diamonds (above the line) denote districts spending above the estimated average and dots (below the line) denote districts spending below the estimated average, for any given test score.

The technique of *multiple* linear regression is conceptually identical, except that it adds more variables with which to predict spending. The additional variables cannot be depicted graphically in two dimensions, but the idea of adding variables to an equation is straightforward:

$$\begin{aligned} \text{spending}_{it} = & \beta_0 + \beta_1 \cdot \text{performance}_{it} + \beta_2 \cdot (\text{teacher salaries})_{it} \\ & + \beta_3 \cdot (\% \text{FRL})_{it} + \dots + \beta_n \cdot (\text{prop val})_{it} + \dots + u_{it} \end{aligned} \quad (1)$$

Spending in district  $i$  and year  $t$  is specified to depend on student performance, teacher salaries (as the key input price), % FRL, other demographic and school variables (such as school size), and a set of “efficiency controls” such as property values (discussed in the next section). The unexplained component,  $u_{it}$ , is the error term representing factors not captured by the measured attributes. It can be positive or negative but has an average value of zero. The regression estimates the coefficients  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ , and so on, to provide the best fit to the data, minimizing the variation in the estimated error term.<sup>8</sup>

The key point here is that the resulting equation is a *spending equation*, which gives an estimate of *average* spending for a district of given performance and other characteristics. There is nothing controversial about this statement; the cost-function practitioners would agree, as this is only the first step in their estimation of the cost, or *minimum* spending necessary to produce a given level of performance.

We defer discussion of the key coefficient on performance,  $\beta_1$ , to a later section, but some of the other coefficients are readily interpreted. The estimated coefficient  $\beta_3$  represents the additional spending, on average, among districts with higher percentages of FRL students, holding other variables constant. In essence it indicates what districts with different levels of poverty are spending. It does *not* represent the extra *cost* required to achieve any given performance level for FRL students. All a positive  $\beta_3$  coefficient in equation (1) would reflect is a tendency of either the state or the district to spend more heavily when there is a greater proportion of

---

<sup>8</sup>In the interest of simplicity, the text omits a number of technical details. For example, these equations are often estimated in logarithmic form for the dependent variable and some independent variables. Also, typically the estimation uses instrumental variables for the performance variable (and perhaps others, such as teacher salaries), as is discussed in a later section.

students in poverty, whereas any similar tendency to spend less on poor students would yield a negative coefficient. This interpretation of  $\beta_3$  holds regardless of whether extra spending is required to increase performance or is effective at doing so.<sup>9</sup>

The distinction is quite important, because coefficients estimated from such equations are regularly adduced to specify cost premiums (or student weights) in school funding formulas (Duncombe & Yinger, 2005). For example, in Missouri, the estimate of  $\beta_3$  was taken to mean that a student receiving a subsidized lunch in an average district is more than 50% “more expensive than a student not receiving a subsidized lunch to bring up to the same performance level,” an interpretation that goes beyond what is warranted from a spending equation (Duncombe, 2007, p. 24).

The interpretation of demographic coefficients is further illustrated by variables for race. As an example, the estimate by Baker (2006b) of the extra spending for Black students in Missouri was 70%.<sup>10</sup> The direct interpretation of this equation is that Missouri spends more on average in districts with higher concentrations of Black students (controlling for FRL, etc.). This is consistent with Missouri’s history of mandated remedies in prior desegregation cases. But because these estimates are drawn from *spending* equations (not cost equations), it is an over-interpretation to conclude that these coefficients represent the required extra *cost* for Black students to achieve any given level of performance.

Consider next the control for teacher salaries. The idea here, drawn from the theory of competitive markets, is that if important input prices are beyond the producer’s control, they are an independent determinant of cost. For such input markets, the producer is said to be a “price-taker.” However, it is highly questionable whether such conditions are reasonably satisfied by teacher markets. Although much of the variation in teacher salaries across districts is correlated with the wages of nonteaching college graduates in the region (labor market), within regions districts vary meaningfully in salaries they pay to teachers. This within region variation draws into question the “price-taking” assumption.<sup>11</sup>

---

<sup>9</sup>To be sure, it is uncontroversial that higher FRL is associated with lower district performance, but the statistical evidence that extra spending systematically raises performance over the observed range is highly controversial. Student-level data from Missouri indicates no relationship between spending and performance of African American FRL students (Podgursky, Smith, & Springer, 2007, Figure 10).

<sup>10</sup>Because of the specific functional form, the estimate varies modestly depending on the percentage of students that are Black. The estimate just given is for the average district in the state, whereas for St. Louis, the figure is 85% (Baker, 2006b). The estimate in Duncombe (2007) also implies a substantial premium, but because of the way that race entered the equation (interacted with FRL) the interpretation is less straightforward.

<sup>11</sup>The collective bargaining environment is a textbook case of the violation of the competitive price-taking assumption for inputs, as the impersonal forces of the market are replaced by relative bargaining power.



Consequently, district variation in teacher salaries likely includes discretionary variation, not simply variation in cost. This problem is recognized by some of the cost-function practitioners, and their attempted solution is discussed in a later section (on instrumental variables).<sup>12</sup> The point here is that input pricing illustrates the difficulty in adapting cost-function estimation from competitive markets to the very different environment of public education.

To see the effect of all the controls in equation (1) taken together, consider each district's *fitted* value for spending. This is the value for each district using the estimated  $\beta$ s in equation (1) and setting the error term to zero. It represents the estimated spending for the *average* district of that *specific* district's characteristics. In the simple case of Figure 2, where performance was the only right-hand-side variable, the *fitted* values are represented by the line and the *actual* values are represented by the diamonds and dots. The difference between *actual* spending and *fitted* spending is the distance from the diamonds and dots to the line (also known as the *residual*).

How is this affected by the addition of all the explanatory variables in equation (1)?<sup>13</sup> The answer is seen in Figure 3. The deviation of actual spending from fitted spending—the amount that each district differs from the regression line—is depicted on the vertical axis, plotted against the district's performance. In effect, Figure 3 replicates Figure 2 except that instead of plotting actual spending on the vertical axis, it plots spending adjusted for performance and other district characteristics including student poverty, race, teacher salaries, and so on.

For St. Louis, the effect of these controls is striking. In Figure 2, St. Louis was the highest spending among districts with comparable test scores. In Figure 3, St. Louis is among the lowest spending of these districts, after *controlling* for district characteristics. St. Louis is a very large district that has high percentages of FRL and Black students that go along with its high spending. Thus, after adjusting for these other factors, Figure 3 indicates that St. Louis spends a bit below (but quite close to) the average of what would be predicted based on Missouri spending patterns.

St. Louis is far from alone in spending below the estimated average of comparable districts: Approximately half the districts in the state fall in the same category, as Figure 3 shows. This is true by definition of averages; because Lake Wobegon

<sup>12</sup>Some practitioners (including Baker, 2006b) use regional cost indexes instead of teacher salaries. This avoids some of the difficulties previously discussed but may only weakly reflect prices faced by districts.

<sup>13</sup>Duncombe's equation includes performance (instrumented), teacher salaries (instrumented), % FRL, % FRL  $\times$  % black, % SPED, indicator for K-12 district, a set of indicators for district size, property values, district income, state aid relative to income, % college educated, % age 65 or older, % housing units owner occupied, median housing price relative to average property values, and a series of year indicators.

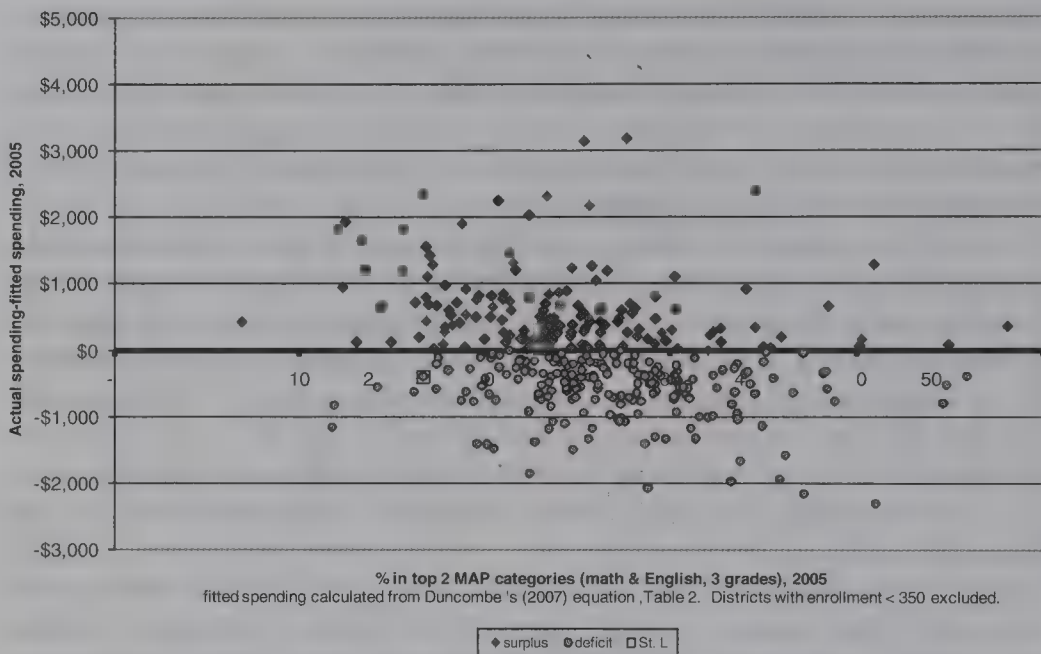


FIGURE 3 Actual versus fitted spending, with controls for district characteristics.  
MAP = Missouri Assessment Program.

is not located in Missouri, about half the districts will be above average and about half below average. The same logic that holds for simple averages carries over to the regression methodology, which estimates average spending among comparable districts. The large number of deficits we saw in Figure 2, for simple regression, appears again in Figure 3—by construction. To interpret these shortfalls from the average as an adequacy shortfall would be logically absurd, as it would mean there is *always* an adequacy shortfall among about half the districts, no matter how high or low spending is. To be sure, these deviations are not quite the adequacy shortfalls implied by the cost function—that requires one further step—but, as we see in the next section, the nature of those shortfalls is largely determined by the deviations shown in Figure 3, results that follow ineluctably from the logic of averages.

Although the statistical controls do not affect the fact that about half the districts spend above and below average, they do affect the size of the deviations. Comparing Figures 2 and 3, we see that the controls help account for some part of the spread in spending over the upper and lower ranges of test scores but did not much reduce the estimated spread in the midrange. The spread containing the bulk of these districts remains about \$2,000 to \$3,000, as it is over much of the test score range. In short, using statistical controls for observable district characteristics helps to identify some spending patterns (e.g., by FRL and race) but still

leaves unexplained a wide range of spending among districts of similar observed characteristics and performance.<sup>14</sup>

### THE ECONOMETRICS OF SPENDING EQUATIONS: CONTROLLING FOR DISTRICT EFFICIENCY

To convert the spending equation to a cost function one needs to identify the minimum expenditure necessary to achieve any given level of performance—the definition of efficient. As Duncombe (2007) pointed out, “Because data is available on spending, not costs, to estimate costs of education requires controlling for differences in school district efficiency” (p. 3).

It is increasingly common to deal with this issue by including “efficiency controls”—variables that are thought to affect efficiency—among the explanatory variables in the spending equation (1).<sup>15</sup> Unfortunately, there is no line item in budgets for “waste, fraud, and abuse.” Moreover, if it were obvious what factors determined inefficiency in schools, local and state citizens and authorities would be likely to take actions to correct the inefficiency. Thus, the quest for a set of observed and measurable factors that convert the spending functions into cost functions by separating inefficiencies from required spending is obviously difficult.

As one example of using efficiency controls, Duncombe’s equation for Missouri includes seven “efficiency-related variables,” categorized as either “fiscal capacity” variables, such as per-pupil property values, income, and state aid, or “monitoring variables,” such as percentage of population 65 years of age or older and percentage of college-educated adults in the district. The argument here is that districts with greater “fiscal capacity” may experience less pressure to be efficient (or a greater inclination to spend on nontested subjects) and that older or college-educated voters may exert greater “monitoring” for efficiency. No analysis—in Duncombe’s report or elsewhere—directly relates any of these variables to efficiency; that is just a maintained hypothesis. In a similar analysis for California

---

<sup>14</sup>This variation could be the result of inadequate controls for true differences across districts. For example, the percentage of FRL students is likely to be a poor measure of the variation in resources that students receive at home across districts, especially across relatively high-poverty districts. Yet these coarse measures are often the only measures available to researchers or to those designing and implementing school funding formulas. However, as previous analyses of achievement show, even with exceptional measures of district characteristics, much of the variation in achievement for districts with the same spending is likely to remain.

<sup>15</sup>Other methods have also been used, which attempt to identify statistically the points at or near the bottom of figures comparable to Figure 3. These methods, stochastic frontier analysis and data envelopment methods, have been used by Duncombe and others in earlier publications (see, e.g., Duncombe, Ruggiero, & Yinger, 1996; Grosskopf, Hayes, Taylor, & Weber, 1997). Recent work, however, including that presented in court, focuses on the method discussed in the text.

districts, Imazeki (2008) included “efficiency controls” but focused on local competition instead of fiscal capacity or monitoring as her measure of efficiency, using the Herfindahl index for the number of districts in the labor market.

These variables are simply added into the spending equation (1). At this point, the equation is still a spending equation—all that has been done here is to single out a subset of the explanatory variables. A district’s age, education, income, property values, the competition it faces, and so on, may well affect spending patterns, over and above the student demographics and other variables, and the estimation of equation (1) sheds further light on those patterns. One may or may not choose to interpret these variables as controls for efficiency (and, if so, they are certainly imprecise controls), but either way equation (1) remains a spending equation.

The typical procedure used to convert equation (1) from a spending equation to a cost function is to standardize the level of efficiency across districts by setting the values of the efficiency variables at uniform levels, rather than the actual district-specific values, and setting the error term to zero as given by equation (2).

$$\begin{aligned} (\text{“cost” of achieving current performance})_{it} = & \beta_0 + \beta_1 \cdot \text{performance}_{it} \\ & + \beta_2 \cdot (\text{teacher salaries})_{it} + \beta_3 \cdot (\% \text{FRL})_{it} + \dots \\ & + \beta_n \cdot (\text{ave prop val}) + \dots \end{aligned} \quad (2)$$

It is common in these cost-function analyses to set the value of the “efficiency controls” (such as property values per pupil) at the statewide average. Setting the error term to zero, of course, is also choosing the average. This means that about half the districts will be found to spend more and half less than the estimated “cost” of achieving at their *current* performance levels. This result is depicted in Figure 4, which presents the difference between each district’s actual spending and the estimated “cost” of achieving its actual performance level.

How are these figures to be interpreted? Spending for a district can be higher than cost because that district may not be using its resources wisely for maximizing the test performance of students. It is, on the other hand, *logically impossible* for a district to spend less than the minimum necessary to achieve actual performance levels. It would be one thing to recognize that “cost” may be imperfectly estimated and there could be a few outliers. But the estimation technique here *systematically* determines that spending is less than “cost” for about half the districts.<sup>16</sup>

---

<sup>16</sup>Cost function analysts acknowledge that they are only estimating “average efficiency,” a term that would seem to modify the definition of cost. However, they continue to state that the estimated cost figures represent what is “necessary” or “required” to achieve any given result, which effectively restores the original definition. Figures 4 and 5 use the “required” terminology, from Duncombe (2007).



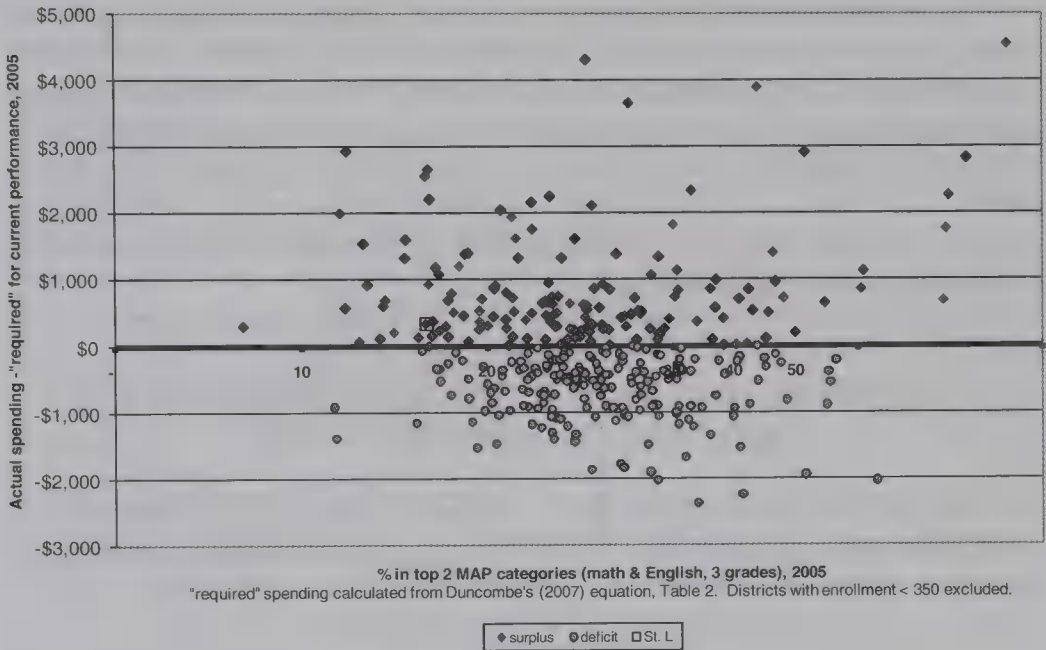


FIGURE 4 Actual versus "required" spending to achieve current performance.  
MAP = Missouri Assessment Program.

Let us be clear on the source of the problem. One might think that the problem is the use of average values for the efficiency variables rather than values that imply something closer to maximum efficiency (minimum spending). This is a valid criticism, but in fact the problem lies deeper.

The primary source of the problem is that the "efficiency controls" do little to explain the variations in spending and are rarely convincing measures of the full range of efficiency. The deviations depicted in Figure 4 have netted out the estimated effect of these variables on efficiency but are still quite large. The step that purportedly converts the spending equation (1) to the "cost" equation (2) has very little effect.

Considering St. Louis as an example, the set of seven "efficiency" variables from Duncombe (2007) taken together tends to raise St. Louis spending above districts with average values of those variables, so the calculated "cost" using those averages is a bit lower than the fitted value in equation (1). Consequently, St. Louis' slight deficit depicted in Figure 3 becomes a slight surplus in Figure 4: St. Louis spends slightly more than is "required" to achieve its actual test scores. As this example illustrates, for most districts there is not much difference between Figures 3 and 4. The interpretation placed on Figure 4 by the cost-function methodology, however, is totally different: cost versus spending. This reinterpretation is not defensible.

In short, the method that purports to convert average spending to cost does nothing of the sort. The adjustment from the “efficiency controls” is minor, not surprising given that it would be difficult to argue that these variables do a good job of measuring true variation in efficiency. The major step is that the deviations depicted in Figure 3—deviations from average spending of comparable districts—are essentially redefined as deviations from “cost.” That is why the “cost” estimates carry the logically incoherent implication that half the districts spend less than is necessary to achieve what they have achieved.

### EXTRAPOLATING FROM THE “COST FUNCTION” TO A DIFFERENT PERFORMANCE LEVEL

The third step in calculating the cost of adequacy is to apply the estimated cost function to a target performance level. This step is accomplished by simply replacing actual performance with target performance in the calculation:

$$\begin{aligned}
 (\text{“cost” of achieving target performance})_{it} = & \beta_0 + \beta_1 \cdot (\text{target performance}) \\
 & + \beta_2 \cdot (\text{teacher salaries})_{it} + \beta_3 \cdot (\%FRL)_{it} + \dots \\
 & + \beta_n \cdot (\text{ave prop val}) + \dots
 \end{aligned} \tag{3}$$

For example, one of the targets considered in Missouri was to raise St. Louis from its current level of 16.7 to a level of 26.3.<sup>17</sup> If we apply this target to all districts (not just St. Louis), the result is to raise the “cost” for those districts below 26.3 and to reduce it for those districts above, relative to their estimated cost of current performance, provided that the estimate of  $\beta_1$  is positive. This obviously increases the estimated shortfall from required spending for the former and reduces it for the latter. This imposes a substantial “tilt” on Figure 4, pushing down the points on the left side of the diagram and pushing up the points on the right.

The result is Figure 5, depicting actual spending versus “required” spending to achieve the performance target of 26.3. These estimates redistribute the shortfalls from higher performing districts to lower performing ones.<sup>18</sup> For example, St. Louis was depicted in Figure 4 as spending slightly *above* what was “required” to achieve its current level, but Figure 5 depicts St. Louis as \$1,541 *below* what

<sup>17</sup>Duncombe (2007) identifies this as the Missouri School Improvement Program standard for St. Louis in 2008. This target happened to be near the state average in 2005, of 25.6.

<sup>18</sup>The fact that the process starts from a logically flawed base can still be seen in Figure 5 by examining the large number of “deficit” dots to the right of the vertical line. These are districts that are found to spend less than “required” to meet the standard that they are already meeting.

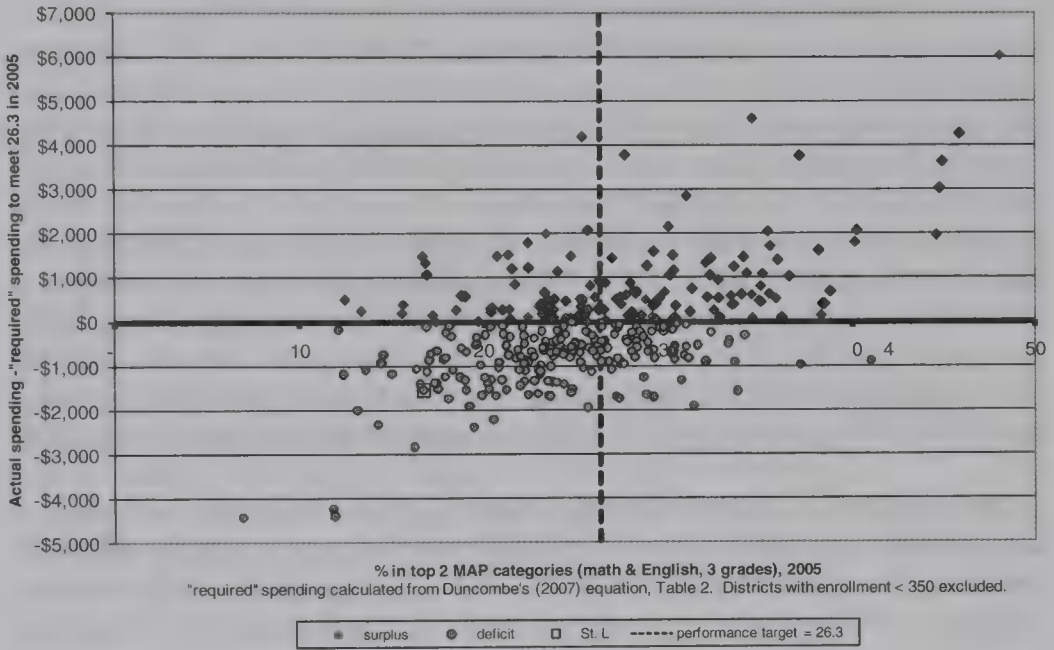


FIGURE 5 Actual versus "required" spending to achieve target performance.  
MAP = Missouri Assessment Program.

is "required" to achieve the higher target. Districts with lower performance are adjusted even further, to yield estimated shortfalls of more than \$4,000.

To assess whether the estimates of cost in Figure 5 are valid, we must directly assess the two key features of the cost estimates: the methodology for estimating the "cost" of generating *current* outcomes, which we have already seen is fundamentally flawed, and the estimated coefficient  $\beta_1$ , which is applied to that base, to generate the "cost" of *target* performance.<sup>19</sup> The estimate of  $\beta_1$  is key to the whole exercise, so it is critically important that it is estimated accurately, with a high degree of confidence, and that it not be sensitive to arbitrary choices in model selection. Unfortunately, there are several reasons why this standard is not met.

<sup>19</sup>Duncombe (2006) and Baker (2006a) have argued that the upward tilt in diagrams such as this in Kansas (Duncombe) and other states (Baker) provide some evidence in support of the approach's statistical validity (albeit a "fairly weak validity test" in Duncombe's view). However, as our step-by-step derivation shows, the tilt simply reflects the estimated sign of  $\beta_1$ . The point is that *any* positive estimate of  $\beta_1$ , even if it is highly problematic (for reasons such as those discussed in the next section), will necessarily generate a positive tilt in a diagram such as Figure 5. Thus, a positive tilt is of no independent value in assessing the validity of the cost-function estimates.

## IMPRECISION IN ESTIMATED COEFFICIENTS, AND IN ESTIMATED "COST"

The first problem is that the regression coefficients are often estimated with relatively wide confidence intervals, even assuming that the model is correctly specified and appropriately estimated. For example, Duncombe's estimate of  $\beta_1$  is that costs rise by 0.39% for every 1% increase in performance. However, the 95% confidence interval ranges from 0.07 to 0.71%, spanning a factor of 10. Similarly, in the study of California districts, the 95% confidence interval for Imazeki's (2008) estimate ranges from 0.05 to 0.63%. Even if everything else is correct, one can have little confidence in the adjustments that lead to estimates of needed costs, moving from Figure 4 to Figure 5, as that depends on the very imprecise estimate of  $\beta_1$ .

The problem of wide confidence intervals applies to the other coefficients as well, which is a matter of some importance for the issue of demographic cost premiums. For example, Duncombe's estimate of  $\beta_3$  implies a premium for FRL students of 52%, but the 95% confidence interval is 27 to 80%. Similarly, there is an implied premium for students in special education of 49%, but the 95% confidence interval is 19 to 80%. Again, these are rather wide confidence intervals, and even they assume everything else is estimated correctly.

The imprecision in all the estimated coefficients, along with the estimated variance in  $u$ , the unexplained component of equation (1), contribute to wide confidence intervals in the estimated "cost" of meeting performance targets. Unfortunately, it is often the case that adequacy estimates drawn from cost functions focus on the point estimate, which may have little value when the confidence interval is wide. For St. Louis, the estimated "cost" of performing at a level of 26.3 in 2005 is \$11,597. However, the 95% confidence interval is from \$8,367 to \$16,074. Because this interval contains the current level of \$10,056, one cannot conclude that spending is inadequate to achieve that target at conventional confidence levels, even if the rest of the analysis is solid. In addition to the problems already identified, however, there are special problems with estimating  $\beta_1$ , to which we now turn.

## SPECIAL PROBLEMS WITH ESTIMATING $\beta_1$ , THE COST OF RAISING PERFORMANCE

There is a long history of trying to estimate the relationship between average spending and average performance, and it is not an encouraging one. For decades, it has proven difficult to find a systematic relationship, and the problems that have plagued that research also pertain to the cost-function estimates. For one thing, the control variables are imperfect, the choice of control variables is arbitrary in some cases, and the estimates are often sensitive to that choice.



More important, it is usually assumed that spending affects performance, as opposed to the opposite that is assumed in the spending relationships that are estimated. Indeed, the whole theory of the court case is precisely that: Providing more resources leads to higher achievement. The implications of this are very serious for the estimation of the spending/cost relationships, because  $\beta_1$  will now reflect both effects *even though just the impacts of achievement on costs are desired*. A related problem is the worry of omitted variables that comes from the possibility of a third factor such as parents' interest in education affecting both spending and achievement. Both of these problems give reason for one to believe that  $\beta_1$  is likely to be estimated with bias.

Cost-function analyses often try to use instrumental variables techniques to reduce bias. However, the requirements of this technique are difficult to fulfill and cost functions to date have not utilized convincing instruments.

Finally, the estimates differ dramatically depending on the specification, whether spending is modeled as a function of achievement or achievement is modeled as a function of spending. We now turn to a discussion of these three problems in estimating  $\beta_1$ .

### *Sensitivity to Selection of Other Variables*

The first problem is that the results are often highly sensitive to which control variables are included in the model. For example, in both the Duncombe and Baker models for Missouri, the results are highly sensitive to the inclusion of race. If race is excluded from the model (as it surely would be, if it were to be used for an actual funding formula), the coefficient on performance,  $\beta_1$  is no longer statistically significant, which is to say the 95% confidence interval includes zero.

Similarly, estimates in Baker (2006b) are highly sensitive to which "efficiency controls" are included in the estimating equation. His data set contains six such variables—similar to those used by Duncombe—though he selects only four of them. Among the 64 possible combinations of those six controls, the  $\beta_1$  estimate is statistically indistinguishable from zero almost half the time, and in most of those cases the model's "fit" is better than the one chosen by Baker. One cannot have much confidence in any single estimate of  $\beta_1$  if both the estimates and the confidence intervals are so highly sensitive to arbitrary choices in model specification.

These sensitivities are found in other states as well. Results provided in Duncombe (2006) show that the estimate of  $\beta_1$  for Kansas loses its statistical significance if an interaction term is omitted (free lunch multiplied by pupil density). If the time period 2000 to 2004 is broken up into 2000–01 and 2003–04, the estimate for  $\beta_1$  doubles between these subperiods, but for neither period is it statistically significant.

*Endogeneity Bias, Omitted Variables Bias, and Instrumental Variables*

A second problem is statistical bias because of mutual causation between spending and achievement (“endogeneity bias”) and/or omitted variables that are likely to affect, or at least be correlated with, both spending and performance. For example, suppose some districts are more education oriented than others, simply because of the gathering of like-minded citizens, with specific characteristics that are not captured by the observable variables. These districts may tend both to spend more and to have more highly performing children. If so, then the relationship between spending and performance will be biased upward, because their statistical association will be picking up in part the effect on each of them of the unobserved degree of education orientation.

The usual solution to this problem is a technique known as “instrumental variables.” Under this technique, “performance” is considered a “troublesome explainer” for spending and does not actually enter into the estimating equation (1).<sup>20</sup> Instead a related variable or set of variables is used, known as “instruments.” The idea is that instead of using variation in achievement that could be a result of a third variable that also affects spending and thus is subject to bias, this technique uses only the variation in achievement that comes from a known source that does not independently affect spending. The theory of this approach is compelling; however, in practice it is rarely well implemented. The problem is that this technique has some stringent requirements, which are rarely met. In the context of cost-function estimation, it is difficult to identify variation in achievement that is the result of factors that do not independently influence spending. If these conditions are not met, the instrumental variable solution to the problem of bias can easily make the problem worse (Murray, 2006).

There are statistical tests that provide some defense against using invalid instruments, and at a minimum the cost functions should pass the relevant test. These tests are weak because they have to assume that some of the instruments are valid to test whether all of them are; yet, in the case of Missouri, the instruments failed these tests for both cost functions submitted to court. Thus, the adequacy estimates were not only methodologically flawed but also statistically invalid.

In addition to the problem of invalid instruments, which lead to biased estimates, there is also a problem of weak instruments—variables that are only weakly correlated with performance. This leads to an overstatement of the statistical significance of the performance coefficient. In other words, the claim that  $\beta_1$ —the key coefficient in the whole exercise—is statistically distinguishable from zero is often undermined by weak instruments. A final difficulty in the instrumental variables approach is that the choice of instruments may be somewhat arbitrary

---

<sup>20</sup>When “teacher salaries” is used as an input price control (as in Duncombe, 2007), it is also treated as a troublesome explainer and instrumented.

and the estimated performance coefficient may be quite sensitive to the choice of instruments.<sup>21</sup>

### *Sensitivity to Specification as “Cost” Versus “Production” Function*

Finally, cost estimates are extremely sensitive to whether spending is modeled as a function of achievement or achievement as a function of cost. There are two traditions looking at the relationship between student performance and spending: the production-function approach and the cost-function approach. The key difference between the two is whether the focus of attention is achievement or spending. Each approach standardizes for a variety of other factors such as economic disadvantage of families, district attributes such as population density, and other things, and then looks at the remaining correlation of spending and achievement. The difference is whether spending is on the left side of equation (1) and performance on the right (cost function) or whether these are reversed (production function).

The first thing to note is that these two approaches must necessarily be related. After all, they look at the relationship between the same basic elements of achievement and spending. Viewing them together provides an easy interpretation of the empirical evidence, but unfortunately this is seldom done. The one exception, where production-function and cost-function approaches are placed side by side, is Imazeki (2008). Imazeki's analysis finds that achieving adequacy in California is estimated to require additional spending of \$1.7 billion if a cost-function estimate is used and \$1.5 trillion if a production-function estimate is used—clearly a striking difference.

Both the cost-function and production-function estimates show weak and imprecise relationships between average district spending and average student achievement, as illustrated in Figure 6 for eighth-grade math scores in the 522 districts of Missouri in 2006. After allowing for differences in the FRL populations, in the racial composition (percentage Black), and in the number of students, one can plot achievement against spending in a way that uses statistical methods to control for the other characteristics mentioned.

Figure 6 shows that there is a slight upward slope of the spending line, but the dominant picture of the figure is (once again) essentially a cloud, where districts with the same spending get wildly different achievement. The line has a statistically significant but relatively small positive slope of 0.0028 scale points per dollar ( $t = 3.1$ ).<sup>22</sup> The flatness of this line is important: Spending more money

---

<sup>21</sup>The Missouri cost functions suffered from both problems discussed in this paragraph, although the point was somewhat moot because the instruments chosen were invalid.

<sup>22</sup>It should be pointed out that Figures 6 and 7 are not necessarily representative of all student outcome measures. If one took a different grade to look at these relationships or looked at reading instead of math, most alternatives actually give insignificant relationships between spending and

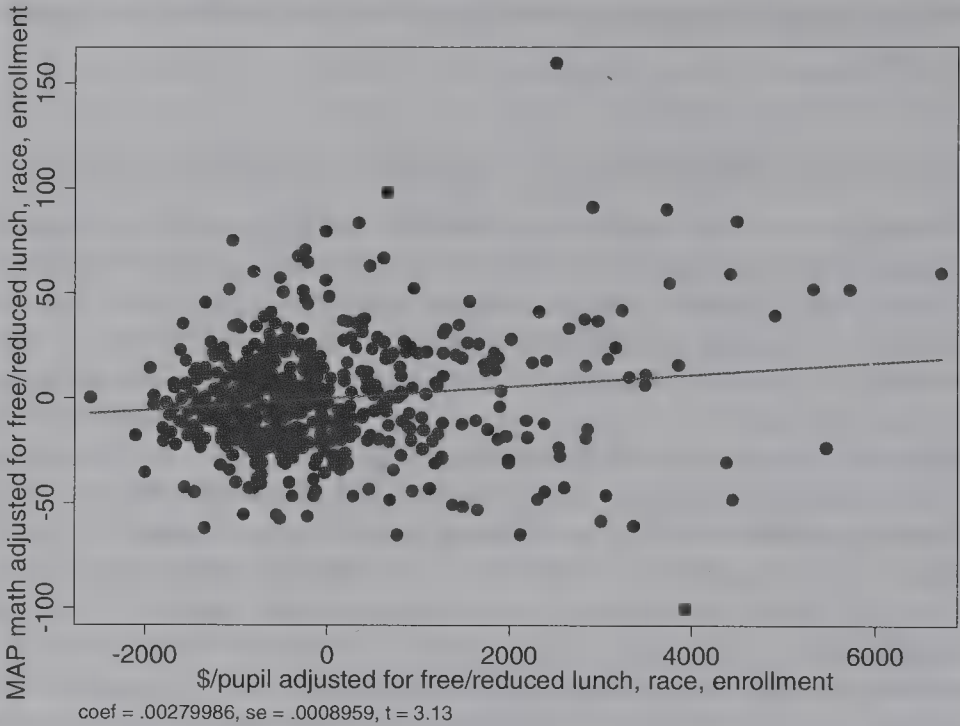


FIGURE 6 “Production function” relationship between eighth-grade math and spending (holding constant race, enrollment, and free or reduced lunch eligibility), Missouri. MAP = Missouri Assessment Program.

given the current way it is spent yields very little achievement gain. Put another way, if one wishes to get a large change in achievement, it will cost a very large amount of money, even assuming that this linear relationship can be extended far away from the current spending: It costs \$357 per pupil to raise achievement 1 point.

Figure 6 is not very encouraging for the proponents of reaching adequate levels of performance through solely spending more money. If it requires tripling or quadrupling funds to get students to the adequate level, most reasonable people will immediately see that this is not a viable public policy.

But there is another way of looking at the data. By looking at how spending varies with achievement—the cost-function approach that we have been discussing—the picture looks far more manageable. Figure 7 turns the previous picture on its side and looks at the amount of spending as a function of

---

achievement, and frequently they have the wrong sign. This might be expected, as the regressions are drawing lines through these clouds of points with little shape to the points that allow estimating such a relationship. A few districts performing at a slightly different point in the cloud can change the slope of the relationship.



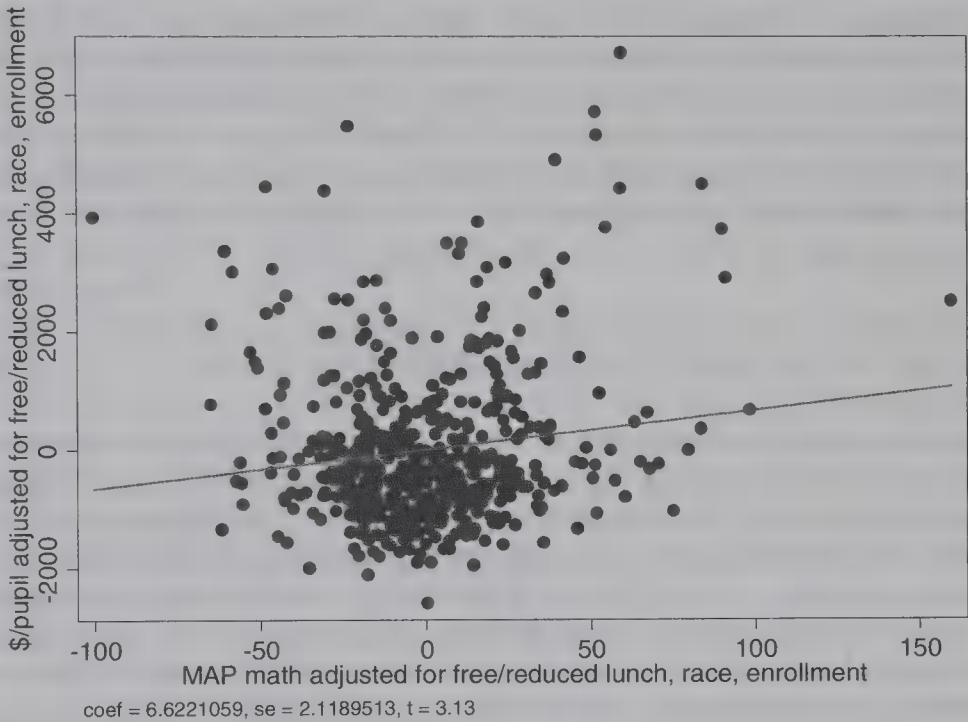


FIGURE 7 “Cost function” relationship between spending and eighth-grade math (holding constant race, enrollment, and free or reduced lunch eligibility), Missouri.  
MAP = Missouri Assessment Program.

achievement (after allowing for the same factors of FRL, race, and district size). Again the dominant feature is the cloud of districts that spend different amounts to reach any given performance level. But now the line that goes through the points tells rather a different story. It is flat once again, but this now indicates that one can move across a vast range of achievement levels at modest cost. The regression coefficient indicates that each \$6.62 raises achievement 1 point.

These regression coefficients reflect the same data—they both have identical  $t$  statistics of 3.13—but they differ dramatically on the estimated cost of raising achievement: \$357 per point versus \$6.62 per point, a factor of 54 (and of course this ignores the wide confidence intervals around each of these estimates). The ultimate reason that these estimates differ so much, even though they use the same data, is that the fit is not very tight. If the fit were perfect, the estimates would coincide: Turning the diagram on its side would turn not only the dots on their side but also the line. However, when the fit is so weak, each diagram will generate a flat curve because they are each minimizing the variation in error terms measured vertically.

The cost function makes it appear that it is much more feasible to change achievement by simply spending more with the current schools and the current

institutional arrangements. For example, in Missouri, the average score on math is 733 and proficiency is defined as 800, so there is a gap of 67 points. Under the cost-function estimate, it “costs”  $67 \times \$6.62 = \$443$  per student to close the gap. Under the production-function estimate, the “cost” is \$23,919. When the estimates vary so wildly from two equally defensible ways of looking at the data—neither one of which finds a strong relationship—it is hard to place much credence on either estimate.

## CONCLUSION

Determining the dollars necessary to provide an adequate education is not an easy task. The commonly employed technique of using professional judgment to design prototype schools is far from satisfying. Case studies of particularly successful schools may provide insights into effective approaches but are also unsatisfying because success is often the function of particularly dynamic leadership or teaching that may be difficult to replicate under current institutional arrangements. Regression-based approaches, often called cost-function analyses, provide a superficially attractive alternative because they apply seemingly objective methods to data on district spending and achievement to determine the cost of reaching standards.

Although on first blush the regression-based approaches are appealing, on further exploration, they are fraught with problems, revealing little about the cost of improving student achievement. The issues facing regression-based models are of two overarching types: technical problems that skilled analysts with sufficient data can correct in their models, and conceptual problems that bring the overall approach into question.

Given sufficient data, a skilled analyst can estimate a regression-based model to produce informative estimates of the spending patterns, by district characteristics and outcomes. Even the most skilled analyst, however, will typically find “cost” estimates that are highly imprecise, sensitive to judgment calls in modeling, and subject to bias.

The underlying difficulty is that even after controlling for a host of variables (including labor market prices, student and school characteristics, among other variables) there is still a great deal of variation across districts in their outcomes for students, in districts with the same expenditures. There are a number of reasons for these differences that draw the regression-based approaches into question. In particular, we have little or no way of knowing how much of this variation is driven by unobserved cost or price differences, by mismanagement, or by a focus on goals other than the student achievement measures used in the cost functions.

Cost-function analysts are aware of these problems. They use efficiency controls and instrumental variables approaches to adjust for these difficulties. However, in

practice both approaches fall woefully short of convincing. We simply do not have good measures of efficiency. The proxies that have been used are, at best, weak measures of efficiency with substantial measurement error, and measurement error itself creates bias. Instrumental variables can, in theory, address the biases due to omitted variables and mutual causation, but in practice no researchers have identified strong and valid instruments. Weak and invalid instruments have been shown repeatedly to overstate statistical significance and to increase bias rather than mitigate it.

The usual practice of identifying “cost” as the average spending among comparable districts always yields the logically impossible result that about half the districts spend less than is required to achieve what they have achieved. This problem has practical implications as well. If courts and policymakers accept a methodology that defines *minimum* expenditures by *averages*, they will then have to raise the expenditures of those below the average, thereby raising the average again. This methodology is a recipe for perpetual findings of inadequacy under forever-recurring litigation.

The failure of regression-based approaches to identify the cost of adequacy is nowhere as clear as when comparing the results of spending as a function of achievement to those of achievement as a function of spending. Cost functions assume that spending changes as a function of achievement; but it makes just as much sense, if not more in the case of education, to assume that achievement changes as a function of spending. A comparison of these two approaches, however, produces vastly different estimates with vastly different implications for policy if interpreted as identifying the causal effect of spending on achievement. Of course, such an interpretation is not warranted.

The cost-function approach simply does not identify the causal relationship between spending and achievement. This failure should not be surprising. We would not need randomized experiments or detailed longitudinal data on student learning to estimate the effects of resources if this could be done so simply with district-level data on spending and average student achievement. However, although not surprising, the problems with regression-based approaches do highlight the difficulty of basing school finance decisions on currently available estimates of the cost of adequacy. All techniques for estimating the cost of adequacy are seriously flawed. None of them can provide a convincing cost figure.

At best, each method provides some limited information—the current distribution of spending and achievement, the cost of a variety of prototype schools, the activities and expenditures in some particularly successful schools. This information can be better than no information for what is ultimately a political decision of how much to spend, but it cannot provide a dollar figure that will guarantee student success or even the opportunity for student success. The most important lesson that emerges from the data—with its wide variation in achievement for comparable expenditures—is that how money is spent is crucial for determining

student outcomes. Educational excellence requires a system with the knowledge, professional capacity, incentives, and accountability that will lead schools to determine how to spend their funds most effectively to raise student achievement and reach the variety of goals we have for students.

## REFERENCES

- Baker, B. D. (2006a). Evaluating the reliability, validity, and usefulness of education cost studies. *Journal of Education Finance*, 32(2), 170–201.
- Baker, B. D. (2006b). *Missouri's school finance formula fails to guarantee equal or minimally adequate educational opportunity to Missouri schoolchildren*. Lawrence: University of Kansas.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., et al. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.
- Costrell, R. M. (2007). The winning defense in Massachusetts. In M. R. West & P. E. Peterson (Eds.), *School money trials: The legal pursuit of educational adequacy* (pp. 278–304). Washington, DC: Brookings Institute.
- Duncombe, W. (2006). Responding to the charge of alchemy: Strategies for evaluating the reliability and validity of costing-out research. *Journal of Education Finance*, 32(2), 137–169.
- Duncombe, W. (2007, January). *Estimating the cost of meeting student performance standards in the St. Louis public schools*. Syracuse, NY: Syracuse University.
- Duncombe, W., Ruggiero, J., & Yinger, J. (1996). Alternative approaches to measuring the cost of education. In H.L. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 327–356). Washington, DC: Brookings.
- Duncombe, W., & Yinger, J. (1997). Why is it so hard to help central city schools? *Journal of Policy Analysis and Management*, 16(1), 85–113.
- Duncombe, W., & Yinger, J. (2000). Financing higher student performance standards: The case of New York State. *Economics of Education Review*, 19(4), 363–386.
- Duncombe, W., & Yinger, J. (2005). How much more does a disadvantaged student cost? *Economics of Education Review*, 24(5), 513–532.
- Duncombe, W., & Yinger, J. (2007). Does school district consolidation cut costs? *Education Finance and Policy*, 3(4), 341–375.
- Gronberg, T. J., Jansen, D. W., Taylor, L. L., & Booker, K. (2004). *School outcomes and school costs: The cost function approach*. College Station: TX, Texas A&M University.
- Grosskopf, S., Hayes, K. J., Taylor, L. L., & Weber, W. L. (1997). Budget-constrained frontier measures of fiscal equality and efficiency in schooling. *Review of Economics and Statistics*, 74(1), 116–124.
- Hanushek, E. A. (2003). The failure of input-based schooling policies. *Economic Journal*, 113(485), F64–F98.
- Hanushek, E. A. (2006). Science violated: Spending projections and the “costing out” of an adequate education. In E. A. Hanushek (Ed.), *Courting failure: How school finance lawsuits exploit judges' good intentions and harm our children* (pp. 257–311). Stanford, CA: Education Next Books.
- Hanushek, E. A. (2007). The alchemy of “costing out” an adequate education. In M. R. West & P. E. Peterson (Eds.), *School money trials: The legal pursuit of educational adequacy* (pp. 77–101). Washington, DC: Brookings Institute.
- Imazeki, J. (2008). Assessing the costs of adequacy in California public schools: a cost function approach. *Education Finance and Policy*, 3(1), 90–108.
- Imazeki, J., & Reschovsky, A. (2004a, July 9). *Estimating the costs of meeting the Texas Educational Accountability Standards*. San Diego, CA: San Diego State University.



- Imazeki, J., & Reschovsky, A. (2004b). School finance reform in Texas: A never-ending story? In J. Yinger (Ed.), *Helping children left behind: State aid and the pursuit of educational equity* (pp. 251–281). Cambridge, MA: MIT Press.
- Ladd, H. F., Chalk, R., & Hansen, J. S. (Eds.). (1999). *Equity and adequacy in education finance: Issues and perspectives*. Washington, DC: National Academies Press.
- Murray, M. P. (2006). Avoiding invalid instruments and coping with weak instruments. *Journal of Economic Perspectives*, 20(4), 111–132.
- Podgursky, M., Smith, J., & Springer, M. G. (2007, October). *A New defendant at the table: An overview of Missouri school finance and recent litigation*. Paper presented at the Show-Me Institute Conference on Education Finance, University of Missouri-Columbia.
- Reschovsky, A., & Imazeki, J. (2003, May). Let no child be left behind: Determining the cost of improving student performance. *Public Finance Review*, 31, 263–290.

# Can Judges Improve Academic Achievement?

Jay P. Greene  
*University of Arkansas*

Julie R. Trivitt  
*Arkansas Tech University*

Over the last 3 decades student achievement has remained essentially unchanged in the United States, but not for a lack of spending. Over the same period a myriad of education reforms have been suggested and per-pupil spending has more than doubled. Since the 1990s the education reform attempts have frequently included judicial decisions to revise state school finance systems. Invoking general clauses about the need for an adequate education found in every state constitution, judges in more than half of the states waded into the development of finely tuned reform strategies. This article empirically estimates the effect of judicial intervention on student achievement using standardized test scores and graduation rates in 48 states from 1992 to 2005. We find no evidence that court-ordered school spending improves student achievement.

The shores of school reform are littered with the wrecks of reform efforts. National, state, and local education leaders have launched an armada of reform initiatives enacted by legislatures or school boards, but none seem to arrive at their destination of school improvement. Perhaps the problem isn't with what reforms are being tried but with who is at the helm. Perhaps judges, who are insulated from electoral pressures, are better positioned than political leaders to identify the circumstances and strategies for effective school reform.

This, at least implicitly, is the rationale for a wave of judicial activity since the 1990s in revising state school finance systems. Invoking general clauses about the need for an adequate education found in every state constitution, judges in more than half of the states waded into the development of finely tuned reform strategies.

Judges heard and incorporated into their thinking claims about the optimal number of students in classes, the appropriate level of compensation for teachers, the ideal school and district size, and a host of other issues that were factored into determining the expenditures that judges would order state legislatures to make. To be sure, legislators had deliberated over these issues on a regular basis, but, the argument went, they had arrived at the wrong conclusions. They were too influenced by re-election pressures and parochial concerns to properly weigh the merits and ensure an adequate education. We needed judges to do the job properly.

Faith in the superior wisdom of judges is not entirely without basis. The most salient example of when judges saved us from the failure of legislatures is the civil rights movement. In that case democracy failed us, perpetuating an obviously unjust and unwise policy of racial segregation. Judges rescued us from that abyss, for which they accumulated a reservoir of popular goodwill. Drawing on that political capital, judges have been empowered to venture into other policy arenas, including education reform. It is not obvious that the intervention of judges in education reform will be as beneficial as the intervention in civil rights. Civil rights is primarily an issue of justice, a question of political values—something at which judges normally excel. Education reform, on the other hand, involves resolving complicated technical questions—something for which judges and judicial procedures are not particularly well suited.

Whether judicial involvement in revising school finance has been beneficial is an empirical question that can be addressed with evidence. The purpose of this article is to assemble, analyze, and present evidence to resolve this question. Have judges succeeded at improving student achievement where others have failed?

To answer this question we examine the effect of judicial intervention in school finance systems on student achievement as measured by test scores on the U.S. Department of Education's National Assessment of Educational Progress as well as by high school graduation rates. We find no effects of judicial action on these measures of student achievement. That is, we find no evidence to suggest that student learning improves as a result of court-ordered changes in school finance systems.

## PREVIOUS RESEARCH

Our results are consistent with the bulk of prior research on related issues. Previous research has generally found little or no benefit for student achievement from adding financial resources to the existing public school system. We can observe the limited usefulness of increasing educational expenditures as a mechanism for improving student achievement simply by examining the temporal relationship between school spending and student outcomes (Greene, 2005). Over the last

3 decades, per-pupil spending in U.S. public education has more than doubled and yet student achievement has remained essentially unchanged.

In 1970–71 public schools nationwide spent per pupil a total of \$4,860, adjusted for inflation to the equivalent of 2005–06 dollars. By 2003–04 that amount had increased to \$10,286 (U.S. Department of Education [USDE], 2006d). Yet during this period outcomes for students showed no significant improvement. According to the USDE's National Assessment of Educational Achievement, the average reading score for 17-year-olds, on a scale from 100 to 500, was 285 in 1971 and was still 285 in 2004 (USDE, 2006b). Math scores show the same basic pattern: In 1973, the average scale score for 17-year-olds was 304, and in 2004 it was 307, a difference that is not statistically significant (USDE, 2006c). According to the USDE's estimate of the long-term trend in public high school graduation rates, 78% of students graduated in the class of 1971 compared to 74.3% in 2004 (USDE, 2006a). Despite more than doubling financial resources for public education, multiple measures of student outcomes show that investment, in aggregate, yielded no gains for student achievement.

Of course, it is possible that other developments negated any benefit that additional spending could have produced. Perhaps the challenges that schools face have increased so that holding student achievement steady is actually a significant accomplishment. Were it not for the additional resources provided to public schools it is possible that student achievement would have experienced a substantial decline.

Although plausible, this claim is at odds with the general trends in factors that affect the challenges students bring to schools. According to the Teachability Index, which tracks 16 indicators, students are coming to the educational process with fewer disadvantages than they were 3 decades ago (Greene & Forster, 2004). In some respects, students have become more challenging to educate. For example, more students come from homes with single parents and where English is not the primary language spoken. But in other respects, students pose fewer challenges; more come to school having attended preschool, the educational attainment of parents has increased, the real income of families (even poor families) has improved, and students have better health. Overall, it appears that students may be easier to educate than they were 3 decades ago. At the very least, it would be hard to demonstrate that conditions have deteriorated so much that they have completely offset the doubling in per-pupil spending.

To isolate the independent effect that additional resources have on student outcomes, researchers have conducted statistical analyses of variation in spending controlling for other observed factors affecting student achievement. Although these analyses can improve upon the internal validity of broad national comparisons of spending and achievement over time, they come at some expense to external validity. If spending had beneficial effects in the particular situation examined, why haven't national measures of student outcomes budged as spending



has doubled? The answer appears to be that although there are isolated studies that are often invoked to prove the desirability of spending increases, the vast majority of econometric analyses of the relationship between school spending and student achievement find no significant relationship between the two. Writing in the *Handbook of the Economics of Education*, Hanushek (2006, p. 25) reviewed the research in this area and finds that 72% of analyses find no statistically significant relationship between student:teacher ratios and achievement, 73% find no relationship between teacher salary and student achievement, 66% find no relationship between per-pupil spending and achievement, and 86% find no relationship between school facilities and student achievement. From this Hanushek concluded, "A wide range of analyses indicate that overall resource policies have not led to discernible improvements in student performance" (p. 38).

Whether increased school spending contributes to higher student achievement is precisely the issue in dispute in school finance lawsuits. But the existence of this dispute in the legal arena does not necessarily mean the issue is in serious dispute in the social science arena. The techniques that are used to justify higher spending levels, including cost-functions, professional judgment models, the "evidence-based" approach, and the successful school models, are more commonly found in the courtroom than in scholarly publications. Refutation of the validity of these techniques has been ably done in previous work as well as elsewhere in this issue and are not be repeated here.<sup>1</sup> It is sufficient to say that the weight of the social science evidence suggests that adding financial resources to the existing public school system should have little or no effect on student performance.

If increased spending has little or no effect on achievement, then it would seem impossible for court-ordered spending to have much effect. But perhaps judges can better identify the circumstances under which additional spending might be more productive and have focused their rulings on those circumstances. In the face of null findings on the general relationship between resources and achievement, the common (and tautological) refrain is that money spent wisely will have a different effect. Perhaps judges know better than legislatures how and when to increase spending so that court-ordered spending will have a different effect from increased spending generally.

For judicial intervention in school finances to affect student achievement, we would probably have to see that intervention resulting in significant changes in school spending. Unfortunately, the current research on this matter suggests that court action actually results in little change in school finances. Just because courts issue orders does not mean that policies will be substantially changed. Legislatures sometimes defy or subvert judicial orders, and sometimes judges order policies that legislatures were going to adopt anyway.

---

<sup>1</sup> See, for example, West and Peterson (2007) and the article in this issue by Costrell, Loeb, and Hanushek.

Earlier research on this issue found that court intervention reduced within-state inequality in per-pupil spending. Murray, Evans, and Schwab (1988) found that equity in funding lawsuits reduced inequality in spending by 19 to 34% between 1972 and 1992. Card and Payne (2002) similarly found that when school finance systems are struck down by courts, the variation in per-pupil spending within states is reduced. They also found that court involvement produces a modest reduction in the variation in student SAT scores, but they did not report an effect on average achievement. Baicker and Gordon (2004) found that school finance judgments increase state aid to local education systems, but that is partially offset by reductions in local education spending and reductions in state aid to localities for noneducational purposes. Springer, Liu, and Guthrie (2005) attempted to disentangle the effects of court intervention based on equity concerns versus those based on adequacy concerns, only to discover that there does not appear to be much of a difference between the two in how school finance is affected.

Berry (2007) updated the data set examined by Springer, Liu, and Guthrie, who in turn updated the data set used by Card and Payne and by Baicker and Gordon. Berry also made an important methodological improvement on the earlier work by using state-clustered standard errors. Berry argued that the standard errors used by earlier work failed to account for serial correlation, inflating the statistical significance of the reported findings. Berry essentially replicated the earlier work but found that with state-clustered standard errors, court action seems to have little or no effect on school finances.

Total education revenue is not significantly higher in states after judges overturn the school finance system. It also does not seem to make a difference whether the courts acted on equity or adequacy concerns. In one model that counts the number of years since a court ruling, Berry (2007) produced an estimate that spending increases by \$30 per pupil per year following judicial intervention, but that estimate just falls short of conventional standards for statistical significance. Berry concluded, "Across a wide range of fiscal outcomes measuring both the level and distribution of education spending, the analysis presented here generally reveals substantively small and statistically insignificant effects of school finance judgments" (p. 233).

The previous research suggests that increased school spending has little or no effect on student achievement, and judicial action has little or no effect on the level of school spending. Given these findings it would be extraordinary to find that court involvement in school finance had any effect on student achievement. But this is precisely the issue we are examining. Perhaps judicial action has subtle but important influence over the composition of school spending that nevertheless results in improved student outcomes despite the gloomy expectations derived from previous research.

## DATA AND RESEARCH DESIGN

Our analysis closely follows the data and research design employed by Berry. We examine whether school finance litigation affects student outcomes using a state fixed-effects model. Berry provided us with a copy of his data set from which we obtained information on judicial actions and state demographic information. We supplemented those data to include updated information through 2005. Data regarding school finance litigation between 2003 and 2005 were obtained from the National Access Network Web site maintained by Teachers College at Columbia University ([http://www.schoolfunding.info/states/state\\_by\\_state.php3](http://www.schoolfunding.info/states/state_by_state.php3)).

The only significant change we make to Berry's data or analytical approach is to replace his school spending dependent variables with student achievement dependent variables. In particular, our dependent variables were state average test scores, standard errors<sup>2</sup> of state test scores, and high school graduation rates. The tests were fourth- and eighth-grade reading and math scale scores on the USDE's National Assessment of Educational Progress (NAEP). The measures of high school graduation rates were estimates produced by the Manhattan Institute (Greene & Winters, 2005, 2006). Obviously, the average test score is a measure of overall student performance. However, most funding lawsuits are initiated to benefit students who are performing poorly, so it is possible that school finance rulings will improve the performance of students at the bottom tail of the distribution without having a measurable effect on the mean. We include the standard error of the distribution to test for this possibility. (Although, if the standard error decreases without changing the mean, it implies potential harm to students in the upper tail of the distribution and could represent an adverse consequence of the finance judgment.) It is also possible that the distribution is unchanged but the improvement has been in preventing dropouts. If the graduation rate increases while the score distribution is unchanged, this could also represent an academic improvement resulting from the school finance judgment.

Because we do not have these student outcome measures before 1990 our analysis also differs from Berry's in that it includes only this more recent period. We have high school graduation rates for each state for each year between 1991 and

---

<sup>2</sup>Ideally we would use the standard deviation of the population, but we had access only to standard errors with NAEP data. Note that the standard error is the standard deviation divided by the square root of the sample size, and we know the population and percentage of the population consisting of school-age kids. For robustness we created a quasi-standard deviation by assuming kids are evenly distributed among the grades within a state and that NAEP samples the same percentage of kids in a state from year to year. This quasi-measure will be the actual standard deviation divided by the square root of the sampling percentage. Because all the models we estimate include state fixed effects, the square root of the sampling percentage will disappear as long as it is "fixed" for each state. When we ran the analyses using this quasi-measure we found qualitatively similar results. For transparency, we report the coefficient estimates when using the standard errors as the dependent variable.

2003. The NAEP data are less comprehensive, with only some states taking the test each year before 2002, when the administration of the test became universal. To accommodate this irregular schedule with our fixed-effects design we divided the NAEP scores into six “eras” or periods during which tests were taken rather than into annual intervals. Doing so measures time slightly less precisely but prevents missing data and should have little substantive effect on our findings.

Like Berry, our unit of analysis is the state and our model has a dummy for each state and for each year (or era). This design allows us to isolate changes in the dependent variable, if any, that occur after court action. Following Berry’s example, we also include in our model controls for some state demographic factors, including the state population older than 65, state school-age population, total population, and income. Finally, we report state-clustered robust standard errors, as Berry suggested.

Our independent variables of interest are measures of judicial action in school finance lawsuits. We have dummy variables that represent whether the courts have overturned the state’s school finance system on adequacy grounds, whether they have overturned that system on equity grounds, and whether they have upheld the funding system. These variables allow us to measure whether student outcomes are different after these court actions than they were before. But it is also possible that the impact of judicial involvement in school finance takes time to yield benefits for student achievement. To capture that possibility we have an alternative specification of the model in which we add a variable that counts the number of years since courts struck down the state school funding system.

In total we present 27 analyses—nine measures of student outcomes with three model specifications for each outcome. The nine dependent variables are state average scale scores and standard error of scores on the fourth-grade reading and math tests, the eighth-grade reading and math tests, and high school graduation rates. The three model specifications are as follows: The first model includes dummy variables for school finance formulas being overturned or upheld (with the excluded category being no challenge thus far), the second model replaces the overturned indicator with dummies specifying whether the school finance ruling was based on adequacy or equity grounds, and the third model adds to these dummies a counter for years since the court overturned the school finance system.

## RESULTS

The clear conclusion across all analyses is that we find no evidence that judicial involvement in state school finance systems improves student achievement. As can be seen in Tables 1 through 6, none of the independent variables of interest



TABLE 1  
Descriptive Statistics

Summary Statistics	1992		2005	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>Court Intervention</i>				
SFJ	0.2500	0.4376	0.5098	0.5049
Equity	0.2174	0.4170	0.2600	0.4431
Adequacy	0.0435	0.2062	0.4800	0.5047
Upheld	0.2800	0.4536	0.4200	0.4986
Years Since SFJ	3.2000	6.5403	8.3137	11.2508
<i>Outcomes</i>				
Grade 4 reading	215.2893	8.3068	218.1141	7.4846
<i>SE</i>	1.2701	0.2663	1.0855	0.2280
Grade 4 math	218.4101	8.2994	237.0534	6.6979
<i>SE</i>	1.1610	0.245	0.8558	0.1768
Grade 8 reading	260.4437	7.1423	261.6131	7.0763
<i>SE</i>	1.2952	0.2568	1.0171	0.2264
Grade 8 math	266.3603	10.208	277.7520	8.5717
<i>SE</i>	1.1838	0.2897	0.9972	0.2477
Graduation rate	0.755	0.0762	0.7229	0.0846

*Note.* The earliest Grade 8 reading scores are from 1998, and the most recent graduation rates are from 2003 data. SFJ = school finance judgments.

is consistently positively related to our measures of student outcomes. When we measure academic achievement using test scores, the upheld indicator consistently has a negative sign, but it is only significant in 2 of 12 analyses and only at the nonconventional  $p < .1$  level. The estimated coefficient for the adequacy indicator is positive and significant only once in 12 equations, and again it is only significant at the  $p < .1$  level, which is likely because of chance when this many equations are estimated.

When the standard errors of exam scores are used, we have similar null findings. In these 12 analyses we estimate 36 coefficients on variables of interest and find only 2 statistically significant at the  $p < .1$  level.

When graduation rates are used to measure academic achievement we find more statistically significant results; however, the coefficients on both the equity and adequacy indicators are *negative*, implying school finance judgments harm graduation rates rather than improve them. Although it is possible that schools somehow altered their priorities, leading to lower graduation rates following school finance judgments, we find it more plausible that these are spurious results. We find no evidence that judicial intervention in school finance leads to improved student achievement.

TABLE 2  
Effect on 4th Grade Reading Scores and Distribution

	Grade 4 Reading Scores			Grade 4 Reading SEs		
% population > 65	1.1160 (0.9400)	1.2580 (0.9600)	1.202 (0.960)	-0.0079 (0.0570)	-0.0165 (0.0570)	-0.0166 (0.058)
% population 5-17	0.3650 (0.4100)	0.2460 (0.4000)	0.224 (0.390)	0.0134 (0.0340)	0.0229 (0.0340)	0.0228 (0.034)
Per-capita income	-3.233** (1.3500)	-3.560** (1.4600)	-3.654** (1.450)	0.1040 (0.1000)	0.1250 (0.1100)	0.125 (0.110)
Per-capita income sq	0.0417*** (0.0160)	0.0460*** (0.0170)	0.0477*** (0.017)	-0.0011 (0.0011)	-0.0014 (0.0012)	-0.00135 (0.001)
Population	0.00253** (0.0012)	0.00246** (0.0012)	0.00239** (0.001)	-0.0002 (0.0001)	-0.0002 (0.0001)	-0.00016 (0.000)
Population squared	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.000)
SFJ indicator	-0.5370 (0.8300)			-0.0155 (0.0570)		
Upheld indicator	-0.9090 (0.8200)	-1.0650 (0.8300)	-1.185 (0.850)	-0.0569 (0.0550)	-0.0620 (0.0560)	-0.0622 (0.056)
Adequacy indicator		-0.9360 (0.8100)	-0.371 (0.880)		-0.0279 (0.0510)	-0.027 (0.057)
Equity indicator		1.7800 (1.3700)	2.219 (1.350)		-0.0820 (0.0960)	-0.0813 (0.098)
Years since SFJ			-0.158 (0.120)			-0.00026 (0.007)
Constant	238.5*** (26.7000)	244.0*** (27.9000)	246.9*** (27.800)	-0.0374 (2.1800)	-0.4420 (2.2700)	-0.437 (2.280)
Observations	206	202	202	206	202	202
No. of states	48	48	48	48	48	48
R <sup>2</sup>	0.49	0.50	0.5	0.38	0.39	0.39
F probability	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note. Robust standard errors in parentheses. SFJ = school finance judgments.  
\*\*  $p < .05$ . \*\*\*  $p < .01$ .

DISCUSSION

In some ways these null findings are completely unsurprising. Given previous research suggesting little or no effect of increased school spending on student achievement and little or no effect of judicial involvement on total school spending, it is unremarkable that we find no relationship between court-ordered spending and educational outcomes. It would have been quite unusual if we had found any other result.

Yet, viewed in another way, these null findings are very unexpected. In more than half of the states, courts have ventured into school finance on the premise that

TABLE 3  
Effect on 4th Grade Math Scores and Distribution

	Grade 4 Math Scores			Grade 4 Math SEs		
% population > 65	0.2300 (1.0400)	0.1330 (1.1400)	0.0836 (1.150)	-0.0374 (0.0570)	-0.0229 (0.0590)	-0.0205 (0.059)
% population 5-17	1.001* (0.5400)	1.015* (0.5300)	0.938* (0.530)	-0.0235 (0.0370)	-0.0245 (0.0360)	-0.0207 (0.036)
Per-capita income	2.4400 (1.5700)	2.741* (1.6200)	2.523 (1.590)	0.0095 (0.0820)	0.0354 (0.0850)	0.0464 (0.086)
Per-capita income sq	-0.0301 (0.0190)	-0.0341* (0.0200)	-0.0311 (0.020)	-0.0001 (0.0010)	-0.0003 (0.0010)	-0.00048 (0.001)
Population	0.0018 (0.0015)	0.0019 (0.0015)	0.0018 (0.002)	0.0001 (0.0001)	0.0001 (0.0001)	0.0000 (0.000)
Population squared	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.000)	0.0000 (0.0000)	0.0000* (0.0000)	0.0000* (0.000)
SFJ indicator	0.0972 (1.1200)			0.0028 (0.0820)		
Upheld indicator	-1.872* (1.0400)	-1.1910 (1.0100)	-1.314 (1.050)	0.0397 (0.0790)	0.0060 (0.0790)	0.0123 (0.080)
Adequacy indicator		1.6910 (1.2800)	2.342* (1.270)		-0.0592 (0.0540)	-0.0921 (0.064)
Equity indicator		0.2310 (1.5900)	0.78 (1.640)		-0.1850 (0.1200)	-0.212* (0.120)
Years since SFJ			-0.147 (0.120)			0.00742 (0.007)
Constant	147.0*** (38.2000)	141.8*** (39.5000)	147.9*** (38.900)	1.6140 (2.0100)	1.0090 (1.9200)	0.699 (1.960)
Observations	127	125	125	127	125	125
No. of states	48	48	48	48	48	48
R <sup>2</sup>	0.92	0.92	0.92	0.62	0.63	0.63
F probability	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note. Robust standard errors in parentheses. SFJ = school finance judgments.

\*  $p < .1$ . \*\*\*  $p < .01$ .

they could alter student outcomes. Tens of millions of dollars have been spent on the litigation. Courts have ordered tens of billions of dollars in increased spending (although if we believe Berry's results, this resulted in little or no more spending than legislatures would have done anyway). Judicial activity has raised serious concerns about separation of powers. The integrity and credibility of the judicial system was put on the line. If all of this was done for naught, that would be shocking indeed.

Unfortunately the evidence consistently shows that judicial involvement in school spending has yielded no improvements in student outcomes. Judges appear to have no special wisdom or advantage over their elected colleagues in

TABLE 4  
Effect on 8th Grade Reading Scores and Distribution

	Grade 8 Reading Scores			Grade 8 Reading SEs		
% population > 65	1.7930 (2.3400)	1.7900 (2.3500)	1.9080 (2.3500)	−0.1430 (0.1200)	−0.1490 (0.1100)	−0.1430 (0.1200)
% population 5–17	−0.7770 (0.7700)	−0.8170 (0.7900)	−0.7060 (0.7900)	0.158** (0.0640)	0.153** (0.0650)	0.159** (0.0650)
Per-capita income	−5.7490 (3.7800)	−5.5360 (3.9100)	−5.5510 (3.8800)	0.2960 (0.2200)	0.3590 (0.2300)	0.3580 (0.2300)
Per-capita income squared	0.0784 (0.0500)	0.0750 (0.0520)	0.0764 (0.0520)	−0.0042 (0.0030)	−0.0052 (0.0031)	−0.0051 (0.0031)
Population	−0.0017 (0.0037)	−0.0018 (0.0038)	−0.0019 (0.0037)	0.0002 (0.0002)	0.0002 (0.0002)	0.0002 (0.0002)
Population squared	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
SFJ indicator	−0.1080 (0.7800)			−0.0867* (0.0510)		
Upheld indicator	−1.1640 (1.0000)	−1.2920 (1.1500)	−1.0680 (1.1700)	−0.0235 (0.0960)	0.0175 (0.1100)	0.0302 (0.1100)
Adequacy indicator		−0.2570 (1.2200)	−0.0142 (1.2400)		0.0558 (0.0880)	0.0697 (0.0890)
Equity indicator		0.0000 0.0000	0.0000 0.0000		0.0000 0.0000	0.0000 0.0000
Years since SFJ			−0.1780 (0.2500)			−0.0102 (0.0160)
Constant	364.8*** (68.5000)	362.9*** (72.1000)	360.4*** (70.6000)	−5.8260 (4.3700)	−6.8930 (4.5300)	−7.0330 (4.5600)
Observations	123	120	120	123	120	120
No. of states	48	48	48	48	48	48
R <sup>2</sup>	0.17	0.17	0.18	0.35	0.33	0.34
Fprobability	0.0444	0.0496	0.0797	0.0076	0.0306	0.0170

Note. Robust standard errors in parentheses. SFJ = school finance judgments.

\*  $p < .1$ . \*\*  $p < .05$ . \*\*\*  $p < .01$ .

legislatures or on school boards in identifying the circumstances and manner in which additional spending would produce better education. Education policy is complicated, is highly technical, and involves strong conflicts of values and interests. Although elected legislators and school board members may suffer from parochial and short-term concerns in assessing these issues, courts suffer from other disadvantages. Courts are lacking in the deliberation and electoral accountability that might assist them in determining the credibility of competing claims about education policy. Without debating colleagues, as they do in legislatures, and without having to answer to voters, the unchallenged thinking of judges may lead them into errors.



TABLE 5  
Effect on 8th Grade Math Scores and Distribution

	<i>Grade 8 Math Scores</i>			<i>Grade 8 Math SEs</i>		
% population > 65	0.188 (0.9900)	0.241 (1.0500)	0.21 (1.0500)	0.0259 (0.0550)	0.0355 (0.0590)	0.0391 (0.0590)
% population 5–17	1.076* (0.5600)	1.131** (0.5500)	1.081* (0.5600)	0.0207 (0.0500)	0.0136 (0.0430)	0.0194 (0.0480)
Per-capita income	1.992 (1.5800)	2.285 (1.6300)	2.141 (1.6300)	0.122 (0.1300)	0.126 (0.1200)	0.142 (0.1300)
Per-capita income squared	−0.0209 (0.0180)	−0.0246 (0.0200)	−0.0226 (0.0200)	−0.00147 (0.0016)	−0.00142 (0.0014)	−0.00165 (0.0016)
Population	0.00227 (0.0015)	0.00233 (0.0015)	0.00228 (0.0016)	0.000139* (0.0001)	0.000140* (0.0001)	0.000146* (0.0001)
Population squared	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000* (0.0000)	0.0000* (0.0000)	0.0000* (0.0000)
SFJ indicator	0.187 (1.2900)			0.0229 (0.1200)		
Upheld indicator	−1.923* (1.0500)	−1.679 (1.0800)	−1.736 (1.1100)	0.0262 (0.0760)	0.000213 (0.0790)	0.0068 (0.0780)
Adequacy indicator		0.769 (1.5100)	1.171 (1.5400)		−0.0256 (0.1000)	−0.0717 (0.1400)
Equity indicator		−0.161 (3.0000)	0.172 (3.1300)		0.0335 (0.1800)	−0.00479 (0.1700)
Years since SFJ			−0.0891 (0.1400)			0.0102 (0.0130)
Constant	199.2*** (39.5000)	191.7*** (40.4000)	195.7*** (40.9000)	−2.219 (3.1400)	−2.344 (2.8100)	−2.798 (3.1000)
Observations	126	124	124	126	124	124
No. of states	48	48	48	48	48	48
R <sup>2</sup>	0.80	0.80	0.80	0.50	0.48	0.49
F probability	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note. Robust standard errors in parentheses. SFJ = school finance judgments.

\*  $p < .1$ . \*\*  $p < .05$ . \*\*\*  $p < .01$ .

Democracy has its virtues as well as its defects. On balance we think the virtues are greater. This is why under normal circumstances our system of government is designed to have policy decisions, like the level of education spending, made by democratic bodies, like legislatures. Our frustrating inability to improve educational outcomes over the last several decades has opened the door to more extraordinary arrangements, including judicial involvement in determining the level of education spending. But the solution to our long-standing problems may not be found in who is driving the spending, courts or legislatures, but in what policies shape the education system in which that spending occurs. It may not be

TABLE 6  
Effect on Graduation Rates

	Graduation Rates		
% population > 65	0.0028 (0.0038)	0.0046 (0.0039)	0.0051 (0.0039)
% population 5-17	0.0047 (0.0031)	0.0028 (0.0031)	0.0031 (0.0032)
Per-capita income	-0.0106 (0.0064)	-0.0133** (0.0066)	-0.0121* (0.0066)
Per-capita income squared	0.0001 (0.0001)	0.000135* (0.0001)	0.0001 (0.0001)
Population	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
Population squared	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
SFJ indicator	-0.0061 (0.0075)		
Upheld indicator	-0.0062 (0.0047)	-0.00922** (0.0045)	-0.00833* (0.0045)
Adequacy indicator		-0.0112* (0.0060)	-0.0140** (0.0060)
Equity indicator		-0.0524*** (0.0080)	-.0552*** (0.0084)
Years since SFJ			0.00111* (0.0007)
Constant	0.839*** (0.1500)	0.909*** (0.1500)	0.872*** (0.1500)
Observations	613	589	589
No. of states	48	48	48
R <sup>2</sup>	0.34	0.38	0.38
F probability	0.0000	0.0000	0.0000

Note. Robust standard errors in parentheses. SFJ = school finance judgments.

\*  $p < .1$ . \*\*  $p < .05$ . \*\*\*  $p < .01$ .

how much we spend as much as the incentive system that shapes whether that spending is used wisely.

None of this is meant to suggest that student outcomes cannot be improved or that increased spending could not contribute to those better outcomes. The problem is that the system in which we have spent, whether by judicial fiat or by legislative act, has squandered those additional resources. Unless we think the next wave of court-ordered spending will yield a result different from the last wave, school finance litigation is not a promising avenue for education reform. The solution will have to be found in revising the structure of the school system and that will almost certainly have to be done in legislatures.

## REFERENCES

- Baicker, K., & Gordon, N. (2004). *The effect of mandated state education spending on total local resources* (NBER Working Paper 10701). Cambridge, MA: National Bureau of Economic Research.
- Berry, C. (2007). The impact of school finance judgments on state fiscal policy. In M. R. West & P. E. Peterson (Eds.), *School money trials* (pp. 213–240). Washington, DC: Brookings Institute.
- Card, D., & Payne, A. (2002). School finance reform, the distribution of school spending, and the distribution of student test scores. *Journal of Public Economics*, 88(1), 49–82.
- Greene, J. P. (2005). *Education myths* (pp. 7–19). Lanham, MD: Rowman and Littlefield.
- Greene, J. P., & Forster, G. (2004, September). *The Teachability Index* (Manhattan Institute, Education Working Paper No. 6). New York: Manhattan Institute for Policy Research. Retrieved October 20, 2007, from [http://www.manhattan-institute.org/html/ewp\\_06.htm](http://www.manhattan-institute.org/html/ewp_06.htm)
- Greene, J. P., & Winters, M. A. (2005, February). *Public high school graduation and college-readiness rates: 1991–2002* (Manhattan Institute, Education Working Paper No. 8). New York: Manhattan Institute for Policy Research. Retrieved October 20, 2007, from [http://www.manhattan-institute.org/html/ewp\\_08.htm](http://www.manhattan-institute.org/html/ewp_08.htm)
- Greene, J. P., & Winters, M. A. (2006, April). *Leaving boys behind: Public high school graduation rates* (Manhattan Institute, Civic Report No. 48). New York: Manhattan Institute for Policy Research. Retrieved October 20, 2007, from [http://www.manhattan-institute.org/html/cr\\_48.htm](http://www.manhattan-institute.org/html/cr_48.htm)
- Hanushek, E. (2006). School resources. In E. Hanushek & F. Welch (Eds.), *Handbook of the economics of education* (pp. 865–908). Amsterdam: Elsevier.
- Murray, S., Evans, W., & Schwab, R. (1988, September). Education finance reform and the distribution of education resources. *American Economic Review*, 88(4), 789–812.
- Springer, M., Liu, K., & Guthrie, J. (2005). *The impact of education finance litigation reform on resource distribution* (Working Paper). Nashville, TN: Peabody Center for Education Policy, Vanderbilt University.
- U.S. Department of Education. (2006a). *2006 digest of education statistics* (Table 99). Washington, DC: Author. Retrieved October 20, 2007, from [http://nces.ed.gov/programs/digest/d06/tables/dt06\\_099.asp?referrer=list](http://nces.ed.gov/programs/digest/d06/tables/dt06_099.asp?referrer=list)
- U.S. Department of Education. (2006b). *2006 digest of education statistics* (Table 110). Washington, DC: Author. Retrieved October 20, 2007, from [http://nces.ed.gov/programs/digest/d06/tables/dt06\\_110.asp?referrer=list](http://nces.ed.gov/programs/digest/d06/tables/dt06_110.asp?referrer=list)
- U.S. Department of Education. (2006c). *2006 digest of education statistics* (Table 121). Washington, DC: Author. Retrieved October 20, 2007, from [http://nces.ed.gov/programs/digest/d06/tables/dt06\\_121.asp?referrer=list](http://nces.ed.gov/programs/digest/d06/tables/dt06_121.asp?referrer=list)
- U.S. Department of Education. (2006d). *2006 digest of education statistics* (Table 167). Washington, DC: Author. Retrieved October 20, 2007, from [http://nces.ed.gov/programs/digest/d06/tables/dt06\\_167.asp?referrer=list](http://nces.ed.gov/programs/digest/d06/tables/dt06_167.asp?referrer=list)
- West, M. R., & Peterson, P. E. (Eds.). (2007). *School money trials*. Washington, DC: Brookings Institute.

# Spending Money When It Is Not Clear What Works

Paul T. Hill

*University of Washington*

Public school funding in the United States is not a product of intelligent design. Funding programs have grown willy-nilly based on political entrepreneurship, interest group pressure, and intergovernmental competition. Consequently, now that Americans feel the need to educate all children to high standards, no one knows for sure how money is used or how it might be used more effectively. This article shows that Americans can learn how to make more effective use of the money available for public schools. But to do so, states and localities must keep careful track of how money is spent; how children are taught and by whom; and what programs, schools, and teachers are most and less productive. Foundations should sponsor rigorous development and testing of new instructional programs, and every level of government should permit experimentation with alternative uses of funds, reproduce effective schools and programs, and abandon ineffective ones.

School district and teacher union leaders claim that public schools can and will educate all children effectively, but only if they get more money. Grassroots educators believe it too, as do parent groups (e.g., the PTA and the League of Women Voters) and large numbers of voters.

The claim is best understood as a political statement, made in pursuit of interest groups' constant objective of getting more money for their clients. But is there anything to it? The claim is valid only if, first, schools now use their money so efficiently that no further improvement is possible with current funding and, second, all schools would become more effective if they got more money.

Our current school financing system is an accident of history. Localities first paid for their schools completely. Then states started to pick up part, or all, of the bill for basic instruction. States further complicated funding by creating separate



accounts for instruction, materials, construction and maintenance, transportation, and so on. Then the federal government created specially funded programs for specific groups of children—children in poverty area schools, disabled children, limited English speakers, and other categories. The states followed suit with their own targeted programs called categorical. Then the federal and state government funded special functions like teacher professional development and evaluation.

The result is that there are many funding sources, each with its own narrow goals. The overall amount available for spending on public education in any locality is the sum of many different funding programs. The total can be computed, but nobody controls it or asks whether it is enough or too much. The amounts we spend and the ways we spend them do not derive from analysis of what is needed and what it should cost. Instead, school spending is a result of many small disjointed decisions made by different levels of government, legislative committees, courts, licensing boards, citizens in bond elections, and school boards in collective bargaining agreements.

No legislative body or school board is responsible for deciding how much is needed to produce a given set of outcomes—say, to ensure that every nondisabled child will graduate from high school or every high school graduate can enter a 4-year college without taking remedial courses. Public education has many funders, and each acts on its own rules of thumb. None is directly responsible for the results or able to calibrate spending in light of evidence about need or performance.

Thus state legislatures and school boards decide how much to spend based on estimates of voters' tolerance for taxes. Sponsors of state and federal categorical programs also get as much as they can in legislative negotiations. Nobody has any idea about how much is enough to educate all children effectively, and the fragmentation of programs means no one is responsible for the overall effectiveness of public investment.

Each funding source has its own goals and rules about how and on whom money can be spent. A familiar number, the districtwide average per-pupil expenditure, can be calculated, but no one controls it. Roza, Miller, Swartz, and DeBurgomaster (2005) wrote of one middle-sized district that kept 200,000 separate accounts for all the grants and subgrants it received. The district's superintendent and chief financial officer didn't know where the district's money was; moreover, their estimates of relative spending on elementary versus middle and high schools were wrong.

Within school districts, most spending decisions are made by separate central office units that are responsible for certain functions (e.g., teacher hiring, purchasing materials, testing, teacher training), not for overall school performance. Money then arrives at the school level not as fundable cash but as people, equipment, and programs that emerge from the disjointed central process. School leaders must use resources for purposes designated by funding sources and the central office. With respect to the most vital resource (teachers), school leaders often have no say

in whom they employ; collective bargaining rules about assignments, minutes of student contact, and class sizes virtually eliminate flexibility in work assignment. School leaders have few opportunities to cash in people or other resources and use the money for something else; they must make do with what they have. Making do is especially challenging in schools serving the lowest income children. Many teachers are assigned there after being rejected or passed over by other schools.

In this situation it is extremely hard to judge whether money is used as effectively as it could be. It is extremely difficult for anyone to know how money is used, much less whether different uses of resources are associated with different student learning outcomes. Moreover, the rules forbid many logically possible uses of funds (e.g., to trade off some teacher salaries for technology investments or to employ a few excellent high-paid teachers rather than many low-paid ones). Thus, it is impossible to observe natural variations in practice to distinguish more from less productive uses of funds.

These generalizations apply to all public schools, including the relatively high performing ones. We have no idea whether they are using money as efficiently as they can to produce student learning or whether more funding would lead to higher performance. The fact that spending is subject to so many disjointed requirements means that educators have little incentive to bend the rules and take chances, as it is far more risky for a school or educator to diverge from rules about use of funds than it is to fail academically. Some educators do experiment, but they must keep it hidden; thus good ideas seldom move beyond their source.

Uncertainty about links between spending and outcomes and inability to experiment is especially intolerable in schools that don't now teach children what they need in order to function as adults. These are generally schools serving low-income minority children—particularly African Americans, Spanish speakers, and Native Americans—especially in big cities. We know that schools serving the disadvantaged generally do not produce the outcomes their students need, but we have no reason to believe that they make efficient use of the money spent on them. We also have no reason to think they could use more money except to pay more for the people and equipment they now have.

In this situation there is a great deal of uncertainty about how to serve the disadvantaged well, yet the financing system makes experimentation and imitation of success difficult, and the incentive system makes the search for better methods unnecessary.

How would we turn this around toward a financing system that encouraged experimentation, imitation of success, and abandonment of failure? That is my question. I sketch a financing system premised on uncertainty about what works and built to sustain a continuing search for more effective methods. I show that more funding is neither a sufficient nor a necessary precondition to school improvement. I can't show that additional funding could never help—it might well lead to improvement in some cases—but I show that a great deal of money is

wasted now and that additional funding is likely to be wasted as long as our public education system is structured to spend more for the same people and instructional methods, not to identify or build on more effective instructional methods.

This article's main argument is that the greatest barrier to knowing how to spend money is our lack of a mechanism for developing, testing, and improving methods of instruction. If Americans would acknowledge that they don't know how to educate some groups of children so that they all have the knowledge to function in a modern economy—particularly low-income African Americans and Hispanic immigrants—we would be forced to develop and test alternatives, and continuously replace less effective with more effective schools and programs. Over time, this would certainly result in improvement and greater clarity about trade-offs between spending and results. In the end I suggest how we might both make performance-increasing innovation possible and restructure our public education system so it is capable of continuous improvement.

The article has three parts: first, evidence that we don't know how to provide effective schools for all students; second, evidence that money is not the main barrier to improvement; and third, suggestions about how we can produce the needed knowledge and predispose public education to use it.

## WE DON'T KNOW HOW TO PROVIDE EFFECTIVE SCHOOLS FOR ALL

U.S. public schools are not as good as they could be for anyone, but they are preparing the majority of White, Asian, and minority middle-class students for higher education, though some slip through the cracks. The school performance problem is severe for low-income minority students (especially African Americans in big cities and Hispanic immigrants) who generally do not learn what they need to learn.

The basic facts about school outcomes for low-income and minority students are well known. They are more likely to abandon school before high school graduation (Greene & Winters, 2005), more likely to be denied high school diplomas because they cannot pass state proficiency tests, and more likely to need remediation should they enter college (Jenks & Phillips, 1998). On the national Assessment of Educational Progress, African American high school seniors get about the same reading and mathematics scores as White and Asian eighth graders (Jenks & Phillips, 1998). In general, average tested academic performance of African Americans is consistently a standard deviation below that of Whites and Asians of the same age (Jenks & Phillips, 1998). That means that the average score of an African American student is at about the 34th percentile for White students.

Celio's analysis of the test score gaps in Washington State reveals something even more alarming: As many as one third of African American and Native



American students score below the 10th percentile for all students (see Huggins & Celio, 2004). Students below the 10th percentile include some who have no measurable proficiency.

There are not many minority students above the 85th percentile, but there are far too many to allow anyone to think that race or ethnicity causes low performance. There is no more reason to believe that minority students' test scores are caused by deficits in their nervous systems than there was to believe the speculation current among testing experts during World War I that immigrant Jews were less intelligent than other Americans.

Moreover, it is clear that minority students' test scores rise when they attend schools that teach serious academic content. In the 1970s Coleman showed that Catholic schools attenuate the correlation between race and test scores (see Coleman, Hoffer, & Kilgore, 1982). Several reports in the 1990s found individual schools with unusually high test scores (Thernstrom, & Thernstrom, 2003).

Arguments about racial and cultural determinism will not go away, but they fail the test of parsimony. Americans have not really tried to make schools more effective for urban minority students. Although there are some noble experiments, the vast majority of schools serving these students are frozen in place by rules and get the worst of everything school districts have to offer—they are more tightly regulated, have more rookie teachers, are much more subject to teacher turnover than other schools in the same districts, and are therefore less demanding and less coherent.

These noble experiments—parochial schools serving the poor and schools like KIPP and Cristo Rey—stand outside the public school system and are financed very differently. However, they are extremely hard to reproduce. KIPP and Cristo Rey depend on special combinations of handpicked leaders, teachers working at wages far below their earning potential, and external supporters (in KIPP's case major foundations and in Cristo Rey's case the Jesuit alumni network that finds students part-time jobs).

Inspiring as these programs are, it is hard to claim they represent a general solution to the problem of educating disadvantaged children. Besides being hard to reproduce, they appear to work for only a subset of the disadvantaged student population. KIPP and the parochial students might not handpick students for admission, but they suffer high rates of attrition among African American students. Cristo Rey is built for Hispanic students whose mild manners make them acceptable employees in businesses and law firms owned by Jesuit school alumni.

These programs are also very expensive, counting the value of philanthropic support and donated time. It is not clear whether these programs raise the money needed to expand indefinitely or find ever larger numbers of educators able to work long hours at low pay. There are, moreover, serious questions about whether these schools fully overcome the educational disadvantages their students came with. Parochial high school students are better bets for college admission than students



graduating from public high schools, so colleges seeking minority students recruit them. But the students' SAT scores are still far below White students' scores (Hill, 1994), and their subsequent college experiences can be difficult. These schools are opening up opportunities, but they are not completely closing the achievement gap.

In this I mean to say nothing against these schools. They are great achievements, and they should be reproduced whenever possible. They are a lifeline for families desperate for something better than their neighborhood public schools. But they neither solve all the problems of the children they serve nor can be expanded to serve all the children who need options. Despite the excellent qualities of these schools, the problem of providing effective schools for the disadvantaged remains unsolved.

In school reform circles it is common to hear the statement, "We know how to provide effective schools for all children, but we lack the political will to do it." This statement is usually well meant as an antidote to despair based on racial determinism. But it is not true. Many well-meaning efforts by able people (e.g., New American Schools designs like Atlas Communities schools led by TedSizer, John Comer, and Howard Gardner) accomplished very little. Many promising school models (e.g., San Diego's High Tech High) have worked far better when operated by their inventors than when reproduced under someone else's guidance. Even the parochial schools, often bare bones designs using traditional teaching methods, have been extremely hard to reproduce in the charter school sector.

What we do know about common characteristics of effective schools for the disadvantaged is abstract and evanescent. They have a moral core, coherent curricula and demanding academic standards, strong social cohesion, teachers with intellectual lives, bonds of trust between school and parents, staff members who agree on goals and methods, adults who take responsibility for showing students links among subjects and between school and the adult world, teachers who collaborate to figure out what struggling students need, and so on.

Effective schools have these attributes and ineffective ones do not, but it is not clear how one would build such a school from scratch or how an existing school that lacks these attributes can get them. They can't be learned efficiently out of a book or in a few training sessions. The chemistry that goes into a great school is no easier to reproduce than the subtle bonds that make a great basketball team. It is one thing to say that John Stockton and Karl Malone used the pick and roll, but quite another to reproduce their success.

Thus, we are nowhere close to knowing what it will take to educate all children, including the most disadvantaged, to the point that they are fully prepared for work, higher education, and citizenship. Moreover, even if we knew all the answers for today we would not be sure about what will be needed a generation hence. The economy will change in ways we can't now predict, and so will the requirements for young people to succeed in it.

Worse, we are not in a good position to learn what can work or adapt to new needs. Existing public schools are stuck in place by regulations and contracts, and charter schools generally draw from the public school labor pool for teachers and principals. Charters face little pressure to experiment or innovate because they can prosper simply by offering a more personalized and stable environment (e.g., smaller classes, K-8 or even K-12 in the same school). Private schools similarly compete on the basis of doing the same thing better. None address the fundamental question of how to educate children whom traditional schools, even good ones, don't teach effectively.

### LACK OF MONEY IS NOT THE MAIN BARRIER TO PERFORMANCE

The introduction gives many of the reasons money is not the key factor. We don't know how it is used now, and there is nothing about the structure of public education that puts a premium on efficient use of funds.

Until recently it was impossible to say exactly how public schools used money. Given the lack of any design or intentionality, money had to be used inefficiently, but we could not say exactly how. Roza's recent work cracks open district and school spending patterns in ways that both reveal many gross inefficiencies and show that schools not constrained by modes of public funding spend money differently than do district-run schools. Roza has disregarded district budget documents, which use salary averaging and often charge schools for their pro rata share of centrally delivered services regardless of the amounts of those services schools use. She found the actual staff members assigned to schools, counted real salaries and benefits, and allocated central office service costs according to the amounts delivered to particular schools. Within schools, Roza counted actual staffing costs per course and per pupil.

One of her most striking new findings is that district-run schools spend much less (in some cases less than half as much) per pupil on "core" classes like basic English and mathematics than they do on electives like art and AP courses. This is in part because of larger class sizes in core subjects and the fact that higher paid senior teachers can avoid work-intensive core courses. District leaders have been pressing for greater emphasis on core subjects, in part because their reputations depend on tested student performance on reading and mathematics. But because costs and money flows within schools are invisible, they did not know that actual spending is unrelated to announced priorities.

Roza's has a number of other recent key findings:

Schools whose uses of funds are not regulated in ways public schools are (e.g. charters, magnets, and private schools) spend their money differently: more on

instruction, more on teacher salaries, but for larger numbers of teachers at lower average salaries. They also employ fewer classroom aides and hire specialists (e.g. for art and music) only part time. (Roza, Davis, & Guin, 2007)

School budgets look very different when central office services are considered. Schools with inexperienced staff members and principals (often the schools serving disadvantaged children) get measurably less from the central office and therefore have less money spent on them than other schools. (Roza & McCormick, 2006)

Most districts discourage or close small schools, which are often proposed as better environments for disadvantaged students, on the basis of perceived higher cost. However, when cost of central office services received is factored in, the smallest schools in a district seldom cost more per pupil, and often cost less, than larger schools. (Roza, 2007a)

School districts' uses of money are seldom connected to their announced school improvement strategies. (Roza, 2007a)

Many teacher union contract provisions control the use of a great deal of money. Such provisions as salary increases unrelated to performance, days set aside for professional development, personal and sick days, class size limitations, teachers' aides and more generous health and retirement benefits than those enjoyed by other professionals, cost many districts nearly 1/5 of their budgets. None of these uses of funds has a detectable link to student learning. Some of the extraordinarily costly time off, health, and retirement benefits could, if turned into salary increases for the highest performing teachers and promising newcomers, lead to significant school improvements. (Roza, 2007b)

These findings do not say for sure how money should be used, but they do suggest that money could be used much more effectively.

Other analyses of the links between spending and student outcomes accept current uses of money and try to estimate how much more would be needed to increase school performance. Some estimates of needed increases are based on the opinions of educators, none of whom have succeeded in teaching disadvantaged children all they need to know. Others are based on studies that have shown detectable increases in achievement in districts that adopted particular programs. They assume that the same programs adopted in other districts will have the same effect as in the districts studied—contrary to the experience of all previous efforts to export instructional programs. In many cases the districts that originated effective programs made other changes in policy and resource allocation. These, however, are not specified in simple prescriptions like “reduce class sizes,” “use instructional coaches,” or “increase spending on teacher professional development.”<sup>1</sup>

---

<sup>1</sup> As an example of such broad prescriptions, see Odden, Goetz, and Picus (2007).

Estimates based on expert opinion vary wildly and have little predictive value, as Loeb (2007) showed. Hanushek (2007) recently demonstrated that the effectiveness claims made for particular programs are wildly exaggerated.

Murnane and Levy (1996) provided an excellent example of a simple prescription—reducing class sizes—that does not work unless many other changes not specified in the prescription also happen. In a Texas district that reduced class sizes in 15 schools, positive changes in student achievement were evident in only 2. The 10 schools made no change in teaching methods, other than to give teachers fewer students. The 2 more successful schools transformed teaching, taking advantage of smaller class size to increase direct student–teacher contact and increase feedback on written work. Class size reduction enabled these changes in instruction but did not cause them (Murnane & Levy, 1996).

To this point this analysis has attacked any claims that we know how to educate disadvantaged children effectively. Now I turn to the question of how Americans can put themselves in a position to track the cost and effectiveness of instructional programs, both to improve what is available to all students and to make informed judgments about links between spending and effectiveness.

## HOW TO ADMIT UNCERTAINTY AND GET HIGHER PERFORMANCE

We don't know what works now because we constrain instructional practice within a narrow band of possibilities governed by laws, regulations, and contracts. It is also clear that a great deal of money is spent on things other than instruction. Taken together these facts mean that we can't know how effective schools could be with the money now available. Surely spending greater amounts of money on the same things would waste even more, but we can't say whether, efficiently used, currently available amounts are a little excessive, about right, or not nearly enough.

To know better what works and to make informed trade-offs between expenditures and outcomes we need a very different system, one designed around the expectation that the best methods are unknown but determined to develop, test, and adopt them.

The uncertainty perspective is appropriate because the current structure of public education does not allow enough variation in practice to allow many new ideas to be tried out and does not search for, or capitalize on, innovations. Even if we knew what worked today, and what it cost, we could not be sure what programs or how much money will be needed in the future: We know neither what skills children of the next generation will need nor what forms of instruction technology will make possible.



How can we move toward a system in which we know better how to spend money because innovation is possible, good ideas spread, and less effective practices are replaced with more effective ones?

We can't get the answer through political decision making. As Moe (2003) has shown, politics favors organized interests (e.g., the unions) over disorganized ones (e.g., innovators with ideas that need to be developed and tested) and leads to policies that are hard to change because regulations and bureaucracies are built up to protect them. The politics of education spending answers the question, "How big an appropriation can the supporters of teachers, or vocational education, or computer literacy, swing for [name the interest group] this year?" It does not answer the question, "How much is needed for a student's education?" Group politics leads to a frozen system, not a continuously improving one.

How can we move toward a situation that encourages new ideas about instruction; constantly encourages development and testing; creates avenues for people with new ideas to put them into practice; creates strong incentives for educators and school leaders to search for more effective methods than they now have; and allows children, teachers, and money to shift from less to more productive options?

Determining how much spending on public education is enough is impossible in the absence of a public education system in which funds from all sources can be used flexibly, ineffective activities must be abandoned, and resources can flow to more effective uses. It almost certainly takes more public funding to educate some children than others. However, it also takes less money to run a highly efficient system, where virtually all funds are applied directly to instruction and student services, than an inefficient one, where spending is driven by political and bureaucratic considerations.

What are the necessary elements of a system based on uncertainty and an unending search for better methods? I think there are seven:

- Total transparency about where and on whom funds are spent, what those funds buy, and the true cost of purchases, including salaries.
- Constant tracking of school results (i.e., student outcomes), both prompt (e.g., test results) and long term (e.g., performance at the next level of schooling).
- Thorough student-level analysis of links among funds spent, programs experienced, teacher characteristics, and student outcomes.
- Analysis to identify less and more productive activities, schools, and people.
- Use of analysis results in decisions to abandon or alter unproductive objects of expenditure, or to imitate or reproduce highly effective ones.
- A way of transferring funds and people from less to more effective activities, e.g. an open labor market for teachers and family choice of schools.
- A mechanism for developing and proving new ideas about how to provide more effective instruction, both in general and to particular groups.

Taken together, these attributes would create a demand for demonstrably better schools and methods of instruction, a supply of proven new ideas, and freedom for people and money to move. It will identify effective instructional programs and therefore provide a basis for determining what education should cost.

Our public education system has none of these attributes. Spending and accounting are based on broad categories (e.g., salaries, benefits, capital, transportation), and these are not tracked to the school or student level. Costs are imputed to schools, so that some schools are charged for central services they do not receive and salary cost averaging hides major expenditure differences among schools. Tracking of results is inconsistent and often relies entirely on tests that are at best decent predictors of long-term outcomes. States are only starting to keep student-based records, and only Florida links teacher, student, and school records.

Although states and localities can identify their schools with the highest scores, none has the analytical capacity to assess the net productivity of a school fully controlling for student attributes. Chicago, Oakland, and New York have started closing the schools that have the absolute lowest scores and creating options for families and teachers, but few other districts have followed suit. The same districts (and New Orleans and Philadelphia) have also sought alternative providers for schools and built incubators for new schools. However, critics are right to claim that none of the options created can be considered “proven” (Gill, Zimmer, Christman, & Blanc, 2007).

Our public education system is based on assumptions of certainty: If we can only put enough money and good people into schools, they will work. Groups fight about financial allotments, teacher licensing, class size, and curricular materials, but all claim to know the right answer. Coalitions that prevail in state and local policy fights are sure their solutions will work and therefore see no reason to invest in close tracking of results or to encourage experimentation with alternatives. Thus, the top-line structure of our public education system is hostile to searching analysis, abandonment of existing structures, and creation of alternatives.

Resistance to new ideas discourages the kinds of rigorous research and development (R&D) necessary to create and prove options. This is not disastrous for groups whose schools work reasonably well. But for groups whose schools don’t serve well, it prevents experimentation with new ideas. Although many competent and well-intentioned entities have created charter schools dedicated to the poor, the vast majority draws from conventional generalizations about “good schools.” In general, such schools strive to manage the conventional model well: They hire the best teachers and principals they can, strive to offer coherent instructional programs, and guarantee a safe and studious environment. Few offer new teaching methods, materials, or extensive use of technology. Many of these schools are more effective, but only slightly so, than the district-run schools from which their students came.

In the next section I suggest that to get dramatic improvements in performance, especially new options for children for whom schools currently do not work well, we need an institutional mechanism to generate new instructional models via formal R&D, as well as the other system changes just listed.

### *Practical Steps*

In theory, the arrangement that best meets these criteria is a perfect market. If families could choose schools and teachers could move in an open labor market, the mechanisms needed to transfer people and money from less to more productive schools would be in place. If new providers could arise, and the ones with the most productive approach to instruction could come to dominate the market and make profits accordingly, all the incentives and opportunities for innovation and continuous improvement would be present. Of course we do not have anything like a perfect market in public education and are unlikely to get one.

As Chubb and Moe pointed out, a perfect market is unlikely to arise in a situation in which government controls spending, providers have vastly more information than consumers (and, I might add, no adult's interest perfectly matches that of a child). These conditions are endemic to public education. Even if markets arose in public education, it would be hard to keep courts from intervening on behalf of losers in normal competitive transactions (e.g., families that wanted to attend a school that had lost too many students to survive economically) and educators who felt they had job rights at such schools. We have already seen courts and legislators interfere with normal market processes attached to charter schools.

One could also imagine a centrally managed system with at least some of these elements. Central management could track spending precisely, monitor uses and results, authorize experiments with new ideas, and mandate transfers of funds and people from less to more effective methods. Some businesses at least try to operate in these ways. Intel tries to create innovations that will supplant its own current products in the marketplace and abandons product lines as soon as more productive ones are available. Many firms have abandoned cost allocation formulas in favor of exact tracking of expenditures and outcomes.

However, in government, central leadership is seldom as stable and authoritative as it is in private firms. Historically, politics has introduced constraints that hide real expenditure patterns and costs and protect existing programs from close scrutiny. As Paul Peterson once said to me, "Politics is about hiding things." Public education is now structured to hide resources, avoid scrutiny, and stabilize existing districts and schools both by controlling the movement of students and teachers and by increasing subsidies for schools that lose enrollment. Even if some district leaders move toward transparency and openness to experimentation, their successors, eager to please vocal constituencies like teacher unions, are likely

once again to protect, hide, and insulate people and institutions from performance pressures.

So how do we introduce marketlike elements to public education, at least enough to generate innovation and a constant search for better schools and better methods? We must not only envision a new system but also dispose of the well-funded and politically protected bureaucratic delivery system that now controls the money and owns the loyalty of millions of teachers and families.

Proposals to wave that system out of existence by creating universal vouchers are not working out politically (Moe, 2001). Suburban and middle-class voters fear that vouchers will make the schools in their communities worse by introducing into those schools students whose needs will erode school quality and by forcing some students now in good schools to move to worse ones. A more modest approach—consisting of a competing system of charter schools that will become so effective that it will draw masses of students from public schools—is also struggling with issues of quality and scale. The dominant system defends itself via politics, and the alternatives are not overwhelmingly better.

Bringing market elements into public education is analogous to bringing markets to a post-Communist economy. Some changes are possible right away, but others will fail because people do not know how to behave in a market or because market elements merge in monstrous ways with the existing system.

How do we move toward a public education system based on an assumption of uncertainty and a constant search for something better? What can we do to set up an inexorable movement in the right direction even if we cannot create all desirable changes at once? I suggest the need for three lines of action:

- Investment in R&D to develop and prove more effective instructional systems especially for children for whom schools don't now work.
- Use of charter schools as a means of market entry for new ideas, and as a way to permit field-testing of new instructional systems.
- Creation of a new policy structure that will make it more likely that superior methods will capture the market and force widespread changes in practice.

Taken together these actions can generate possibly more effective instructional methods and forms of schooling, and provide evidence about required amounts and uses of money.

### *Research and Development*

In the early days of the choice movement, research and development was considered a natural byproduct of operating in a competitive environment with the freedom to innovate. However, as we have seen, nothing is automatic. Charters have little capacity to experiment with new approaches to instruction. Even groups



of schools organized into for-profit Education Management Organizations (EMOs) and nonprofit Charter Management Organizations (CMOs) have relied on good management of conventional instruction much more than on innovation. (A partial exception: Edison is reportedly experimenting with uses of online instruction that might reduce the staffing of each of its schools by one or two teachers).

The absence of serious R&D has handicapped the charter movement because it has at best weak proof that it can offer schools that will produce better results than regular public schools. Critics of the No Child Left Behind provisions requiring districts to consider charters as alternatives for children in failing schools get traction when they claim that charter schools might not be any better. In general, the lack of proven methods—and effective mixtures of teacher work, materials, and technologies that can be readily reproduced—is a major deficit in the charter movement.

Imagine how much more open parents and voters would be to charter schools if there were instructional models that lived within available budgets but offered highly effective instruction demonstrated through clinical trials. Such models might look conventional on the surface, combining disciplined teacher work, use of online instructional packages to teach subjects in which teachers are weak, detailed tracking of student progress on key skills, and rapid adaptation of instruction in light of individual student results. However, they might be highly unconventional, for example, delivery of most instruction online with limited use of teachers as diagnosticians and tutors.

Even more significantly, the lack of positive provisions for R&D makes it unlikely that our schools will ever solve the problem of educating children for whom conventional method and organization don't work. This problem is not likely to be solved by entrepreneurship or informal tinkering. It almost certainly requires a formal R&D enterprise analogous to those that develop medical therapies or new defense systems that integrate human and machine work in order to accomplish new missions.

In medicine and defense, new technologies do not simply emerge through tinkering on the part of practitioners. Specialized institutions (e.g., Defense Advanced Research Projects Agency [DARPA] and the National Institutes of Health [NIH]) organize R&D. These institutions work as intermediaries: They fill the space between lab scientists and small-scale inventors on one hand and end users (e.g., physicians, military units) on the other. Thus, for example, DARPA combines previously separate ideas about propulsion, sensors, data processing, structural materials, and aerodynamics into a whole system called an aircraft. It then subjects the new design to extensive tests and uses the test results to put pressure on the armed services to adopt the new system. NIH similarly builds new therapies, essentially systems, out of new discoveries in biochemistry, pharmacokinetics, and delivery devices, then pays for their evaluation and works to get them introduced into practice.

Defense and pharmaceutical industries also conduct R&D, but they depend on the intermediary institutions, and on special indirect cost recovery provisions in federal contracts that set aside money for R&D, to fund the transitions between isolated technologies and whole systems that can be sold and operated.

These intermediaries exist for three reasons: because basic scientists and inventors do not have the financial resources or the knowledge of field operations to build their discoveries into complete systems, because risks are high and failures are necessary, and because the end users have little incentive to try something that disrupts their operating routines or threatens to make cherished skills obsolete. Intermediaries have the funds to assemble multiple emerging technologies into systems, develop the systems, subject them to rigorous proof, and press for adoption of those (systems) that accomplish needed new tasks or do existing tasks better. (Rich's book *Skunk Works: A Personal Memoir of My Years at Lockheed* [Rich & Janos, 1994] is particularly eloquent about resistance to new systems and the need for rigorous testing and demonstration to spur adoption).

Parallels to education are strong. There are many inventors, from individual teachers to software companies, developing small instructional and testing modules. Paul Allen's APEX company has developed online AP courses but not programs for whole schools, and the independent online vendor K-12 targets a niche market: It was unable to enter the vacuum created by the destruction of New Orleans' brick-and-mortar schools. Home schoolers, entities seeking to serve school dropouts, and schools in remote areas have been open to heavy use of online instruction. But school districts have not been interested, except insofar as online instruction attracts children who might otherwise not attend school at all. Districts adopt technology applications piecemeal, as add-ons to existing courses, or for tutoring. School districts and unions resist any approach to instruction that would change the work of teachers or reduce the numbers needed. Entrepreneurs, understanding that there is money to be made in providing what districts want to buy, stick to piecemeal programs or marginal situations.

Thus in K-12 education, innovative ideas exist but they are seldom assembled into whole instructional systems, that is, plans for combining technology with student and teacher work in new ways. Nor are individual technology applications or whole instructional systems tested to the point that their effectiveness and best application can be considered proven. Moreover, there is no mechanism for moving new instructional systems into widespread use.

Earlier efforts to create something like instructional systems—the New American Schools initiative sponsored by businesses in the early 1990s and the federal government's Comprehensive School Reform Program—were not R&D efforts at all. Most designs were put into practice without testing, and few amounted to much more than ways of promoting teacher collaboration. Although many of the designs tried to increase achievement for low-income children, their implementation was so chaotic that it is impossible to say what worked and what didn't. As a result

we know little more about how to create effective instruction for disadvantaged children than before these initiatives started.

The charter sector is particularly handicapped by the lack of proven innovations. Charter leaders, like public school principals and superintendents, can choose from among many plausible but unproven theories about what will work with their students. They have trouble convincing skeptical government agencies that they should be trusted with a school, and once they get a charter, if their first ideas don't work well the only recourse is trial and error.

The intermediary function—finding possible technical applications and new ideas about teacher and student work, assembling groups of them into potential whole instructional systems, subjecting systems for rigorous test, publication of evidence of effectiveness, pressing for adoption, and monitoring field experience for further evidence of effectiveness and limitations—is missing in K-12 education.

A serious R&D initiative to identify promising new instructional systems would cost tens of millions per year, counting development costs of multiple alternative systems and testing costs including controlled trials with real students.

Government has pockets deep enough to pay for development and testing of many alternative systems and to tolerate the inevitably high rate of failure. However, federal government efforts to create NIH-like entities in education (e.g., the National Institute of Education in the 1970s and The Institute for Education Sciences today) have foundered on the same politics of competing certitudes that have rendered the public education itself unable to improve. Government-funded R&D in education is highly responsive to “the field” and tends to celebrate the current conventional wisdom rather than seek alternatives to it. Federal education research agencies have also been unable to overcome educators' resistance to real experiments with tight control of treatments and random assignment of students. Today's government-funded What Works Clearinghouse can identify studies that use strong quasi-experimental methods, but it has no power to create experiments.

I have suggested that major foundations (e.g., Gates and Broad) could also afford to make annual multimillion dollar R&D investments and could make a unique contribution by doing so. They are mulling the possibility, but the lure of continuing to pour money into the next hot idea or hero superintendent is very strong. It is not yet clear whether the mega-foundations will fund this. However, a much smaller philanthropy might fund such an effort if it were willing to concentrate its resources.

When asked to comment on the potential of a major R&D initiative in education, NIH and DARPA experts observe that the “uptake” problem in education is especially severe. School districts and state governments, driven by union politics, prefer to ignore ideas that would change teacher work. Parents want more effective schools, but they can easily be persuaded that smaller class size is the only way to improve instruction.



In contrast, there are strong political advocates for use of high technology in defense and medicine, and the public believes in it. Moreover, there is real evidence of success and failure in both fields, and dramatic consequences of failure to adopt the best methods (though in the defense case there can be many years between disasters that dramatize the need for innovation). In education, however, the providers resist outcome measurement and get away with laying responsibility for school performance on parents, neighborhoods, or society in general.

In this environment how can we move toward innovations that will produce more effective instructional systems and therefore give us real evidence about how much money is needed and how it should be used? Moving the whole public education system at once is too hard, and appealing to only the most innovation-minded teachers and administrators leads to short-term initiatives that are abandoned as soon as a key person tires or takes a new job. The key is to enter public education via the one part of it that has strong incentives to find and use performance-enhancing initiatives—charter schools.

### *Charters as the Tip of the Wedge*

No matter how rigorous it is, an R&D initiative can do only so much. New programs can make a difference only if they are used. This is impossible in a public education system where money is obligated in long-term commitments to people and buildings and where adults are insulated from performance pressure.

Charter schools need to give families and teachers reasons to choose them. In the poorest inner cities, unfortunately, charter schools don't need to be particularly effective to attract families and teachers. They can offer a slightly safer and more studious environment, and that is enough to set them apart.

However, most charter school leaders are dedicated to meeting the needs of a particular set of pupils and would rather be more effective than less. They also face serious cost constraints (e.g., less money than the regular public schools with which they have to compete for teachers) and difficulty keeping teachers. Thus innovations that would make teachers more productive (e.g., uses of technology to deliver information and make linear presentations of material, leaving professionals to diagnose and tutor) could be highly attractive. In addition, hundreds of new charter schools start up every year; even if existing schools found it hard to change their staffs and uses of budgets, new schools are completely flexible and many see an advantage in being known as distinctive and innovative.

The R&D initiative just described could give charter schools the ideas and methods they need to compete effectively. Multischool providers (the EMOs and CMOs previously described) could gain a tremendous advantage if they could adopt proven, reproducible methods. Foundations that sponsor R&D can promote adoption of proven ideas by building new EMOs and CMOs around proven methods.



Charter schools could also serve as sites for full field-testing of instructional systems proven in clinical trials. As in medicine, large-scale use of a system would reveal interactions and consequences too rare to be seen in controlled environments. As in defense, field trials would reveal the need for adjustments in training and support and sharpen estimates of cost. As I have suggested to the major foundations, a final stage of the R&D process, which some intermediary must pay for, is close unbiased tracking of program implementation and student effects.

Today, however, charter schools are secretive and resist analysis. Although they must report good financial data, few create record systems that link student, teacher, and program characteristics to student outcomes. If charter schools are to become laboratories of innovation, they must open themselves in ways I have suggested the rest of public education needs to do. They too need to cooperate with constant tracking of student outcomes; student-level analyses of links among funds spent, programs experienced, teacher characteristics, and student outcomes; and analyses to identify less and more productive activities, schools, and people.

Charter schools are not funded for these activities, and many states exclude charter schools from their testing programs. Legislative action to make sure that all children in the state are tested at state expense and the same records are kept for all students is clearly needed. State budgets also need to include money for collecting information on charter school programs, teachers, and expenditures.

Existing charter schools will not welcome demands for data, though schools built around innovative instructional systems should be more receptive. Any school leader will rightly object that public school systems are staffed for such reporting while they are not. If states are not willing to defray the additional costs of reporting that charter schools must do, philanthropies might need to grant money for design and maintenance of school data systems.

### *Policy Change*

Charter schools would be promising test beds for new R&D-based instructional systems even under existing laws. However, it is important to work toward charter laws and policies that create a level playing field so that charters are not so starved of money or hamstrung by regulation that they cannot compete effectively with district-run schools. A recent Koret task force book, *Charter Schools Against the Odds*, lays out a detailed policy agenda for making charters more effective and for increasing the competitive pressure they exert on district-run schools. Its elements, which overlap with but are not as demanding as those for a whole system based on uncertainty and on unending search for better methods, include the following:

- Equalizing funding for students in charter and traditional public schools via student-based, not program-based, state and local funding systems.
- Empowering new authorizers, including colleges and universities, mayors, and qualified nonprofits in states where school boards hold a monopoly on authorizing charter schools.
- Protecting charter schools from arbitrary denial of applications by establishing appeal processes, to a state agency or independent body, in each state.
- Eliminating arbitrary caps on the numbers of charter schools so that the number of charter schools depends only on the availability of competent and willing school providers.
- Eliminating fixed terms for charter schools in favor of provisions that make it clear a school's charter is valid only as long as it can demonstrate student learning.
- Eliminating bans on for-profit firms holding charters directly, in favor of common funding and oversight provisions for all charter schools, no matter who runs them.
- Allowing charter schools to employ teachers and administrators in whatever numbers, and with whatever mixtures of skills and experience necessary to deliver the school's instructional program. All authorizers have ample power to reject a charter proposal in which the staffing plan does not match the instructional methods to be used.

There is a clear need for a national legislative advocacy agenda, one pressing for needed changes in charter school laws in every state. As *Charter Schools Against the Odds* recommends, a coordinated 50-state agenda, modeled on the Business Roundtable's campaign for standards-based reform in the mid-1990s, could bring about conditions conducive to innovation and competition.

## CONCLUSION

These three initiatives—an R&D intermediary, using charters as the point of the lance, and creation of a level playing field for competition—could set off a wave of innovation and escalating school performance. This, in turn, could tell Americans what they need to spend for effective schools, especially for students who don't now have them, and what higher levels of spending could bring. However, defenders of the existing system will do all they can to disrupt these initiatives, tilting the playing field against innovative schools and fighting the premise that R&D can produce validated, reproducible instructional systems. Unions and schools of education will certainly fight the ideas of clinical trials and random student assignment.

Given the resistance, there is no chance that the whole country, or even a whole state, would adopt all of the policy and funding measures just described. However, some localities (e.g., New York, New Orleans, Chicago) have already adopted student-based funding and other key policies. They might be ideal locales for an effort to use charter schools as the point of the lance for experimentation with new instructional systems. Other localities would have reason to imitate them if they reaped the benefits of improved schooling options for the most disadvantaged and gained greater clarity about productive ways to use public funds.

A well-funded and ambitious R&D intermediary is the one indispensable element that is missing everywhere. If it could be funded and isolated from political interference, it could produce real evidence about what is possible and what it costs. Linked to policy action to level the playing field for charter schools, and to CMOs and EMOs that will put innovations into practice, the products of R&D could include answers to now-unanswerable questions about educational effectiveness and cost.

## REFERENCES

- Chubb, J. E., & Moe, T. M. (1990). *Politics markets and Americas schools*, Washington: Brookings Press.
- Coleman, J. S., Hoffer, T., & Kilgore, S. (1982). *High school achievement: Public, Catholic and private schools compared*. New York: Basic Books.
- Gill, B., Zimmer, R., Christman, J., & Blanc, S. (2007). *State takeover, school restructuring, private management, and student achievement in Philadelphia*. Santa Monica: RAND.
- Greene, J. P., & Winters, M. (2005). *Public high school graduation and college-readiness rates: 1991–2002* (Paper No. 8). New York: Manhattan Institute.
- Hanushek, E. (2007, Summer). The confidence men: Selling adequacy making millions, Stanford. *Education Next*, 3. Available from <http://www.hoover.org/publications/ednext/7273471.html>
- Hill, P. T. (1994). The educational advantages of choice. In T. Moe (Ed.), *Private vouchers* (pp. 120–135). Stanford, CA: Hoover Press.
- Hill, P. T. (ed) (2006). *Charter schools against the odds*. Stanford, CA: Education Next Press 2007.
- Huggins, E., & Celio, M. B. (2004). *Closing the achievement gap in Washington State: Holding schools accountable for equity*. Seattle, WA: Center on Reinventing Public Education.
- Jenks, C., & Phillips, M. (1998). The black–white test score gap. In C. Jencks & M. Phillips (Eds.), *An Introduction* (pp. 1–51). Washington, DC: Brookings Press.
- Loeb, S. (2007, August 6). *Difficulties of estimating the cost of achieving education standards* (SFRP Working Paper No. 23). Seattle, WA: School Finance Redesign Project, University of Washington. Available from <http://www.schoolfinanceredesign.org/pub/workingpapers.shtml>
- Moe, T. M. (2001). *Schools, vouchers, and the American public*. Washington, DC: Brookings Institute.
- Moe, T. M. (2003). The Politics of Status Quo, In Paul Petersons (ed). *Our schools and our future: Are we still at risk?*
- Murnane, R. J., & Levy, F. (1996). Evidence from fifteen schools in Austin, Texas. In G. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success* (pp. 93–96). Washington, DC: Brookings Institute.

- Odden, A., Goetz, M. E., & Picus, L. O. (2007, March 14). *Paying for school finance adequacy with the national average expenditure per pupil* (SFRP Working Paper No. 2). Seattle, WA: School Finance Redesign Project, University of Washington. Available from <http://www.schoolfinanceredesign.org/pub/workingpapers.shtml>
- Rich, B., & Janos, L. (1994). *Skunk works: A personal memoir of my years at Lockheed*. New York: Bay Back.
- Roza, M. (2007a). *Allocation anatomy: How district policies that deploy resources can support (or undermine) district reform strategies* (SFRP Working Paper No. 24). Seattle, WA: School Finance Redesign Project, University of Washington. Available from <http://www.schoolfinanceredesign.org/pub/workingpapers.shtml>
- Roza, M. (2007b). *Frozen assets: Rethinking teacher contracts could free billions for school reform*. Washington, DC: Education Sector.
- Roza, M., Davis, T., & Guin, K. (2007, July 16). *Spending choices and school autonomy: Lessons from Ohio elementary schools* (SFRP Working Paper No. 21). Seattle, WA: School Finance Redesign Project, University of Washington. Available from <http://www.schoolfinanceredesign.org/pub/workingpapers.shtml>
- Roza, M., & McCormick, M. (2006). Where the money goes: District allocation practices are harming high-poverty school. *School Business Affairs*, January, p. 1–4.
- Roza, M., Miller, L. J., Swartz, C., & DeBurgomaster, S. (2005). *Peeling back the layers of funding: An examination of intra-district expenditures*. Seattle: Center on Reinventing Public Education, Daniel J. Evans School of Public Affairs, University of Washington.
- Thernstrom, S., & Thernstrom, A. (2003). *No excuses: Closing the racial gap in learning*. New York: Simon and Schuster.



# Next Needed Steps in the Evolution of American Education Finance and Policy: Attenuating a Judicially Imposed Policy Distraction, Activating a Balanced Portfolio of K-12 School Reforms, Advancing Rationality as a Goal in Pursuing Productivity, Advocating Change in a Responsible and Effective Manner

James W. Guthrie

*Department of Leadership, Policy, and Organizations  
Peabody College of Vanderbilt University*

This article asserts that although there has been a consistently increasing demand on both the national and state levels for alignment of resources (inputs) to improved student outcomes (outputs), the lack of a systematic and well-defined policy portfolio has limited reform effectiveness. This article specifically examines the overreliance on standards and curriculum as reform mechanisms and the often distracting and unproductive judicial interventions connected to equity and adequacy litigation.

Preamble: This polemical article proffers four positions:

- Evolving complexity presently renders education finance policy and education policy generally one in the same. There is no longer a separate disciplinary field or policy specialty of education finance.

---

Peter J. Witham was instrumental in undertaking the research for this paper.

Correspondence should be sent to James W. Guthrie, Vanderbilt University's Peabody College, Peabody #329, 230 Appleton Place, Nashville, TN 37203-5721. E-mail: james.w.guthrie@vanderbilt.edu

- The United States has a reasonably clear K-12 schooling objective, students' elevated academic performance. However, it presently has no clear, consistent, comprehensive, or coordinated policy means for achieving this objective.
- The United States has not financially shortchanged K-12 education resources and long have been plentiful.
- It is possible through concerted collective effort to construct a strategy for aligning resources with the objective of enhanced student performance and thereby having a reasonable chance of improving American education.

### A MODERN PARABLE

Once there existed a powerful, well-intentioned, and wealthy nation. The people and their representatives decided that the nation's children should learn more in school. To achieve this goal, for 50 years the nation continually spent more money on its schools, employed more people to work in the schools, and strove mightily to ensure that these resources were equitably distributed to all schools and all children. The nation even provided more money to schools that educated disabled children or children of poor parents and from poor neighborhoods. The nation also experimented with multiple means for making its schools more effective. Alas, little of this national effort seemed successful. Student achievement in mathematics and reading did not change much, and the gap between middle-class and poor children persisted. What was the powerful and wealthy nation to do?

### INTRODUCTION

During the past half century, America's education finance policy has been bifurcated, blurred, blunted and bloated. Moreover, because education finance policy is now virtually the same as education policy generally, American K-12 education policy suffers from the same lack of purpose. Legislative and executive branch school improvement initiatives compose a crazy quilt of policy options, seldom possessed of a clear focus or representative of a balanced portfolio of reform ideas. A judicial preoccupation with equity has done little to enhance education effectiveness and has bled both policy system energy and financial resources away from crucial issues of school improvement. Given the sustained post-World War II escalation in education spending, and the long stagnant nature of U.S. academic achievement, a refocusing of policy appears in order.

## HISTORICAL PERSPECTIVE

The overwhelming contemporary issue facing education finance is arraying resources so as to propel pupil performance. The shift in the focus of education reform movements (displayed in Table 1) during the second half of the 20th century illustrates this trend. However, for a half century, since the 1960s, the nation's policy system components have been splintered and confused in their pursuit of school effectiveness. Since the 1983 release of *A Nation at Risk*, much of state education reform, and most of federal education policy, has been directed at elevating student achievement. However, state, local, and federal reform portfolios have been unbalanced in favor of a single unproven notion that standards and various kinds of curriculum, instructional, and testing alignment will elevate academic achievement. Other potential powerful reform strategies such as competition and greater reliance on market motivations have been given short shrift. Moreover, regardless of the education reform strategy or strategies involved, there is little by way of a systematic effort to appraise effectiveness and therefore little ability to learn for the future and to undertake mid-course corrections. Finally, the quest for added pupil performance often has been subordinated to a 50-year-long crusade for resource parity. This oft times self-serving equity campaign has done little to improve schools and has served as a policy system distraction. The distraction has been reinforced and rendered persistent by the dominant intervention of the judiciary.

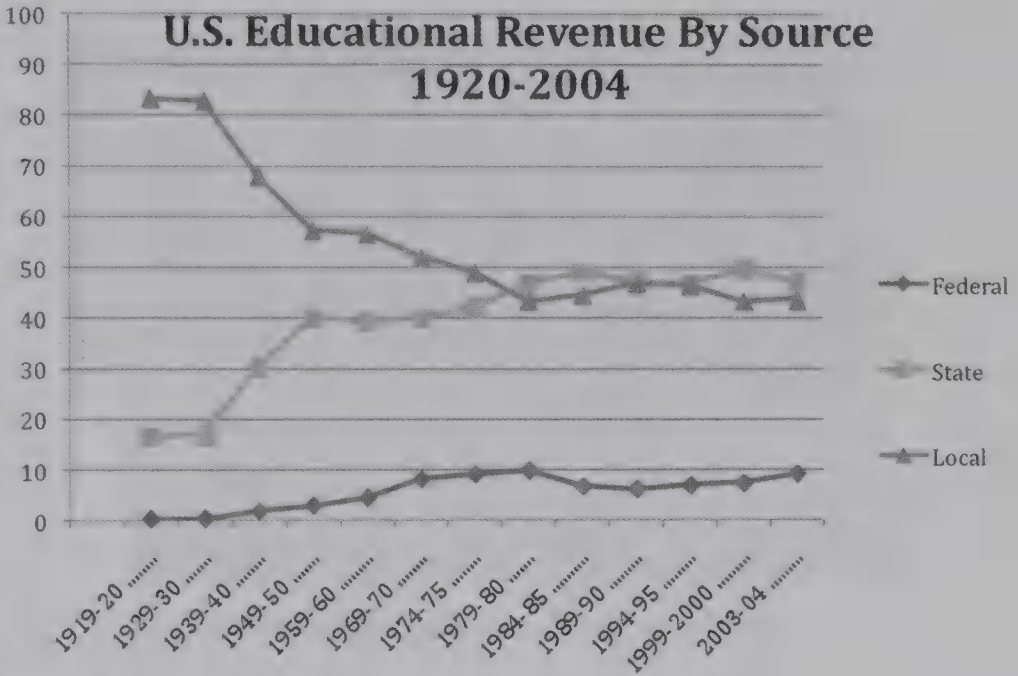
A retrospective view of U.S. education finance over the past half century, 1960 to the present, reveals the following significant conditions.

- Per-pupil expenditures have consistently risen, substantially exceeding costs of living increases.
- Added expenditures have purchased added personnel, not added pupil performance
- State courts have insinuated themselves into the conventional legislative and executive branch policy initiation role, have strongly influenced education finance distribution, and have blunted efforts to elevate achievement.
- State and federal education reform efforts have contributed to a distorted reform agenda, one that privileges a narrow set of technical ideas regarding learning standards, curriculum alignment, and testing at the expense of more venturesome ideas involving market forces such as competition and performance incentives.
- The half-century-long nationwide quest for school finance equality was a consequence of a calculated post-World War II reform activist public policy and media-facilitated campaign.
- No comparably orchestrated or comfortably funded education policy effort has been undertaken on behalf of a coherent policy aimed at creating a full portfolio of reforms or constructing a systematic means of learning from reform experimentation.

TABLE 1  
Major U.S. Educational Reform of the 20th Century

Reform/Condition	Champion	Evidence	Financing	Outcome/Consequences
Scientific management	Business CEOs/ Ed schools/ Media	Anecdotal/ Testimonial	Media/Government	Widespread adoption
Progressive-era depoliticization	Muckrakers/Media	Muckraking/ scandal	Media/Government	Dysfunctional outcomes
School district consolidation	Business CEOs/ Taxpayers/Media	Financial/ Anecdotal	Businesses/ Taxpayer groups	Widespread adoption
Common school movement	Academics	Communitarian Philosophy	School Districts	Dysfunctional outcomes
Racial desegregation	Supreme Court/Civil rights groups	Constitutional/ psychological	NAACP/ Philanthropic	Widespread/Continuing
National Assessment of Education Progress	Academics/ Ralph Tyler	Logical/ Testimonial	Philanthropic	Mixed Results
Education finance equity	Civil rights attorneys/Foundations	Statistical/ Financial	Philanthropic	Dominance of comprehensive schools.
Education standards and alignment	Blue Ribbon Commission	Logical/ Testimonial	Philanthropic/ Government	De jure eliminated De facto accelerated
				Ongoing/Highly regarded
				Successful/ Dysfunctional
				Successful/Incomplete





Source: NCES 2005

FIGURE 1 U.S. educational revenue by source 1930–2005.

### *America's Trajectory of Sustained Added Spending on K-12 Schools*

Individual states are responsible for the statutory provision, combining state and locally generated revenues, of more than 90% of the operating funds for America's public schools. Federal funds compose the overwhelming majority of the remainder (see Figure 1). These funds amount to more than \$3.3 billion each school day.<sup>1</sup> On a daily operating basis, this exceeds the U.S. defense budget.

Figures 2 through 4 display various facets of nationwide K-12 spending over the past century.<sup>2</sup> All figures control for inflation.

From Figure 2, one can see that, even keeping dollars constant, the past century has been a period of almost never-ending upward per-pupil spending. Figure 3 reveals that this ascending pattern is not restricted to any particular kind of

<sup>1</sup>The remaining approximate 10% of operating school revenues stems from federal, philanthropic, charitable, and ad hoc individual payments to school districts.

<sup>2</sup>Figure 2 was assembled by the National Center for Education Statistics (NCES) in 2006. The data stem from U.S. Department of Education, National Center for Education Statistics, Biennial Survey of Education in the United States, 1919–20 through 1955–56; Statistics of State School Systems, 1957–58 through 1969–70; Revenues and Expenditures for Public Elementary and Secondary Education, 1970–71 through 1986–87; The NCES Common Core of Data, "National Public Education Financial Survey," 1987–88 through 2002–03.

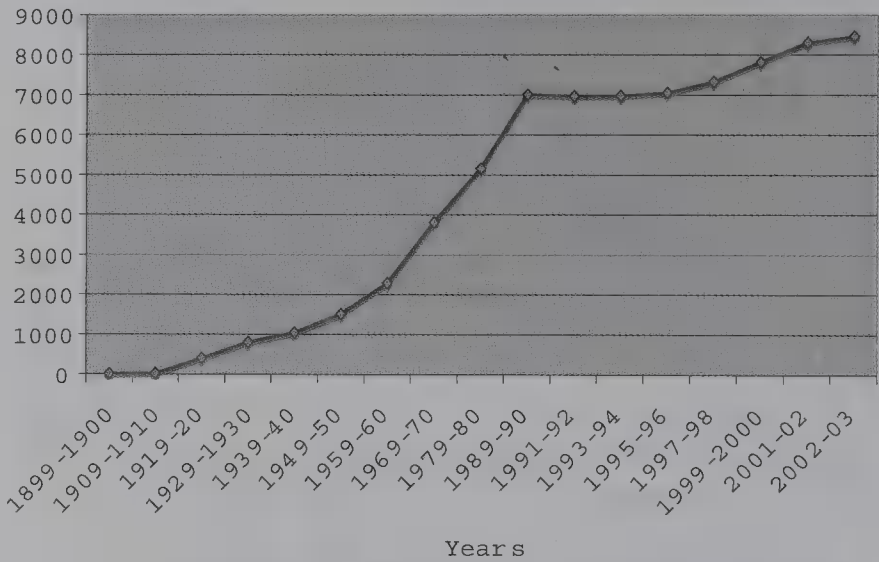


FIGURE 2 Adjusted per-pupil expenditure 1899–2003.

district. All districts, rural urban, and suburban, have been spending more money. It is evident from Figure 4 that the rate of spending increase has particularly accelerated over the past 20 years.

The added amounts of money for schools have been used principally to purchase more labor. Figure 5 displays the ever larger number of employees for America’s public schools. Whereas virtually every other economic sector (e.g.,

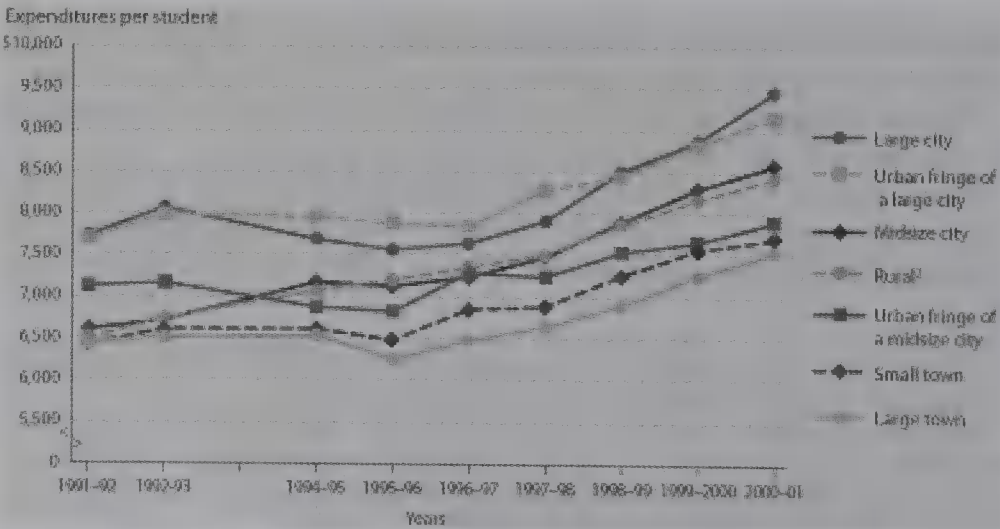
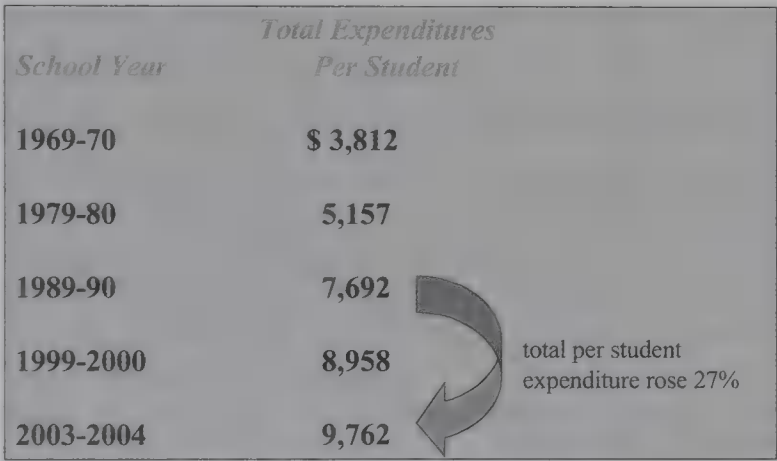


FIGURE 3 Per-pupil expenditures by community type.

**U.S. Per Student Expenditure 1969-2003**  
**K-12 in Constant 2004-2005 Dollars**



Source: NCES, Conditions of Education 2007, p. 75

FIGURE 4 Per-student expenditure 1969–2003.

manufacturing, communication, finance, agriculture, and retail) has been substituting capital for labor, America’s schools have been operating in the opposite direction.

*What Has the Nation Purchased With Its Added K-12 Funding?*

The easy answer is that America’s school districts have purchased more labor and paid existing labor more.

Teacher salaries have increased by 26%, from a constant dollar \$38,665 in 1962 to \$48,165 to 2004 (see Figure 6.) In addition, teacher pensions and other fringe benefits appear steadily to have increased (Costrell & Podgursky, 2008).

In 1995, Rothstein and Miles found that much of the increase in school spending was attributable to added school service features such as the admission of handicapped students to regular schools and classrooms, the serving of meals in school cafeterias, and the transport of pupils. More recently, in addition to some additional services, added school resources have been translated into added numbers of teachers and other employees. The mean teacher–pupil ratio has dropped from 23 students for each teacher 37 years ago to approximately 15 students per teacher today (see Figure 7).

If one posits that smaller classes advantage students and contribute to higher levels of academic achievement, then the added spending is useful. However, the

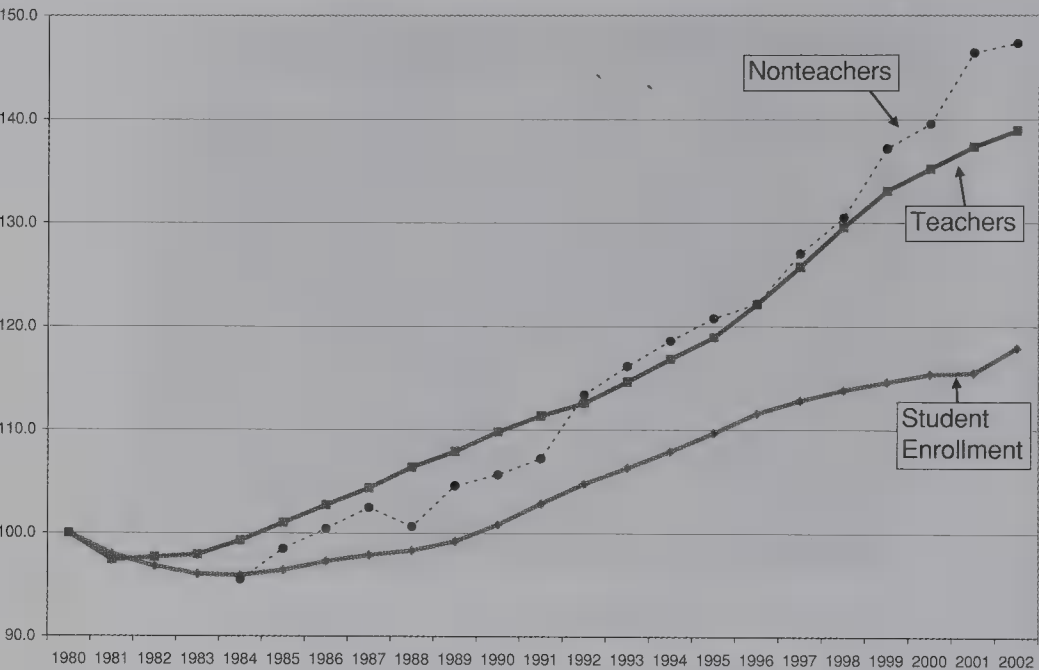


FIGURE 5 Trends in classroom/nonclassroom positions compared to student enrollment (1980–2002).

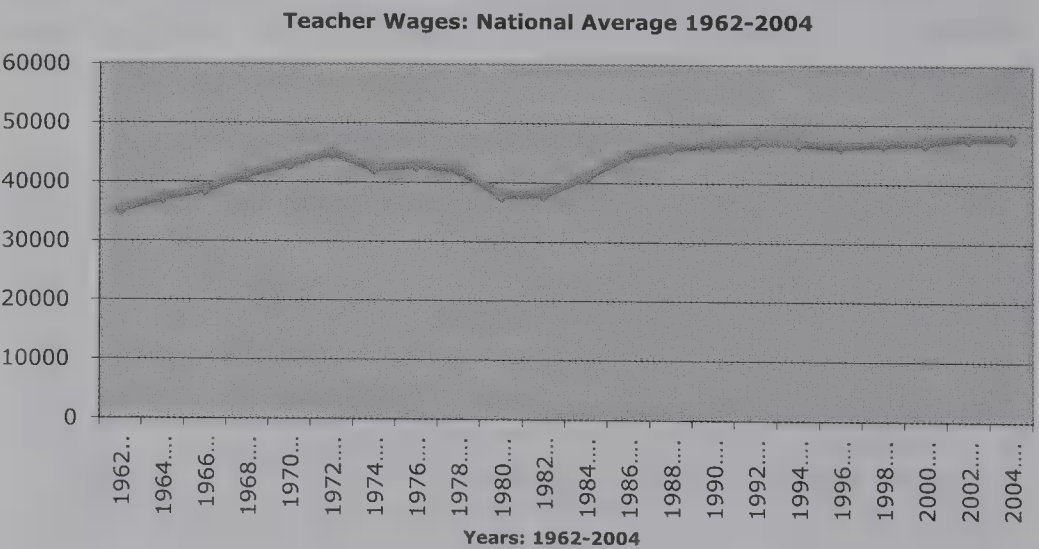


FIGURE 6 National average teacher salary 1962–2004. *Note.* Sources: U.S. Bureau of the Census, Historical Statistics, Colonial Times to 1970, National Center for Education Statistics, Digest of Education Statistics, Bureau of Labor Statistics, Consumer Price Indexes. Bureau of Economic Analysis, GDP and Related Data.



evidence supporting such a proposition is unusually thin and, under the best of circumstances, supports higher academic achievement gains from lower class sizes only in the primary grades (Mosteller, 1995). No reliable evidence supports such a position at the upper grade levels.

### *What Has Been the Return on This Sustained Investment?*

What has the nation received in return from its sustained trajectory of more resources for public school? If one chooses to measure the consequences of spending increases in terms of pupil performance, at the very best, the picture is mixed. The only reliable measure of pupil achievement, available since 1966, is the National Assessment of Education Progress (NAEP). This examination routinely appraises the performance of a national sample of fourth- and eighth-grade students in reading and mathematics. Figure 8 displays mathematics results for 9-, 13-, and 17-year-olds for the years from 1973 until 2004. Here one can see some gains. Seventeen-year-olds' scores have increased only 3 points over the 2 decades, from 301 to 304. However, 9-year-olds' scores increased from 219 to 241, some 22 points. This is approximately a 1% gain in each year over 20 years. During this same period, in constant dollars, school spending averaged a 4% gain each year.

Data in Figure 8 are from a national probability sample of all U.S. students. What about the achievement of subgroups within the overall population? Often a major concern among policymakers is for narrowing the test score differences between White and minority students. If overall achievement was steady, but there were major gains by minority students, then perhaps the added expenditures were justified.

On this topic, the following summary quote by prominent researcher Andrew C. Porter is informative. In a recent article from the University of Pennsylvania Graduate School of Education, titled "Rethinking the Achievement Gap," Porter (2007) commented,

Consider just reading performance among nine year olds from the year 1971 to 1999. The achievement gap did narrow over this period of time into the 1980s, some progress was made, but from that point on, the gap stabilized. The situation is basically similar for mathematics and not so very different for science. (p. 1)

Porter proceeded to point out that not all states are the same when it comes to majority achievement differences. For example, he specified that in Maine, the achievement gap is but one third of a standard deviation, whereas in Wisconsin and Connecticut the gap is much larger, approximating a full standard deviation. An informed reader will immediately recognize that more effective

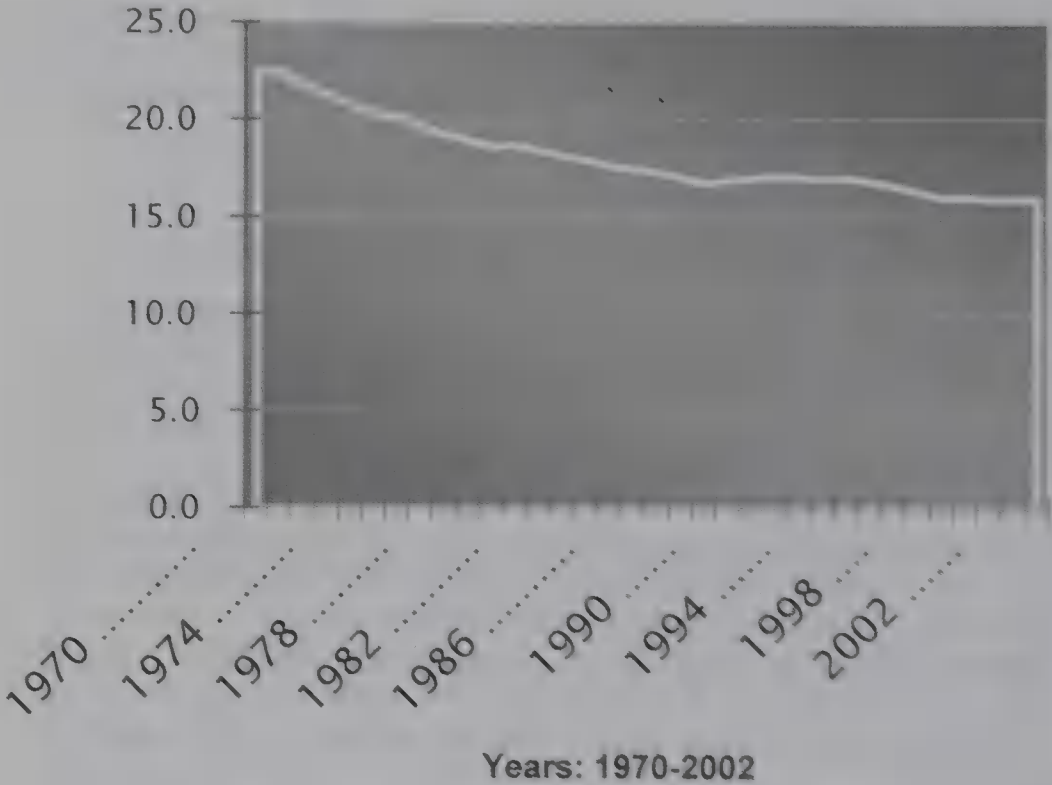


FIGURE 7 National average teacher:pupil ratio, 1970–2002.

Maine is a low-spending state and less successful Wisconsin and Connecticut are high-spending states.

Although NAEP is standardized across the national sample, each state is, nevertheless, free to set its own learning goals, establish its own state testing scheme, and set its own targets for measuring proficient. Figure 9 displays the gap in percentage of students scoring proficient by state standards relative to national standards. Apparently, one means by which states continually can justify increases in school spending is by proclaiming increases in student achievement. The comparisons of percentage of students proficient by state measures with those specified as proficient by NAEP standards suggest that state claims of productivity are often vastly overstated.

Figure 10 reveals that although the average mathematics score of U.S. fourth graders was 518 in both 1995 and 2003, the standing of the U.S. students relative to their peers in 14 other nations was lower in 2003 than in 1995. For example, in 1995, U.S. fourth-grade students were statistically outperformed in mathematics by peers in 4 nations and outperformed peers in 9 nations. In 2003, however, U.S. fourth-grade students were statistically outperformed in mathematics by peers in 7 nations and only outperformed peers in 7 nations.

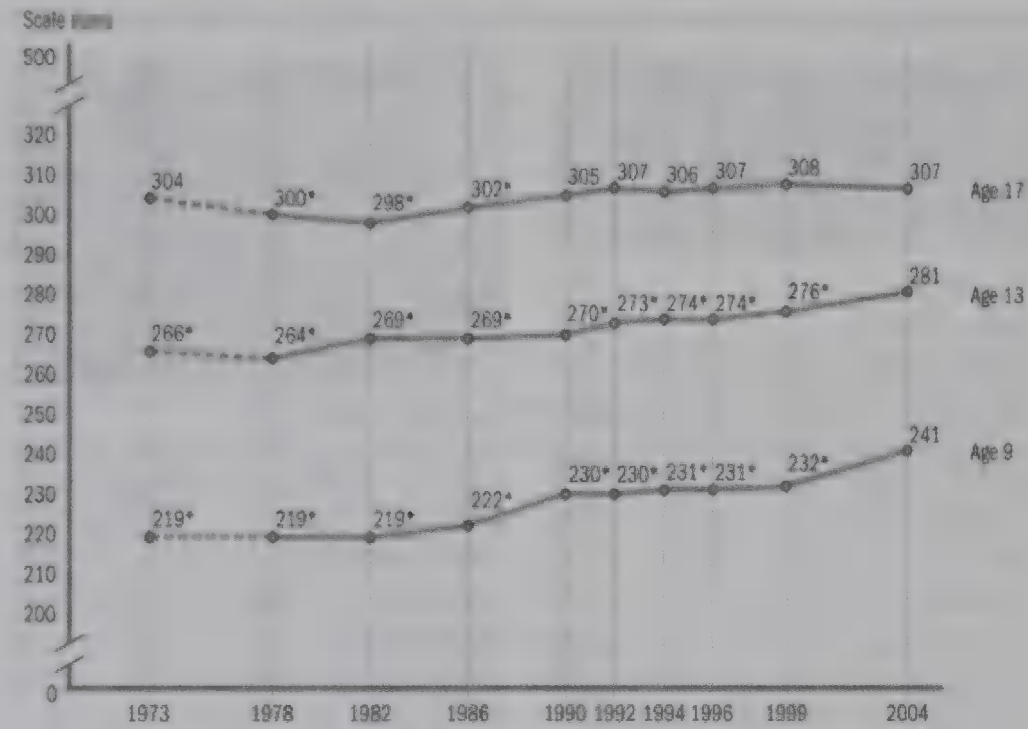


FIGURE 8 Average mathematics scores (9-, 13-, and 17-year-olds), 1973–2004.

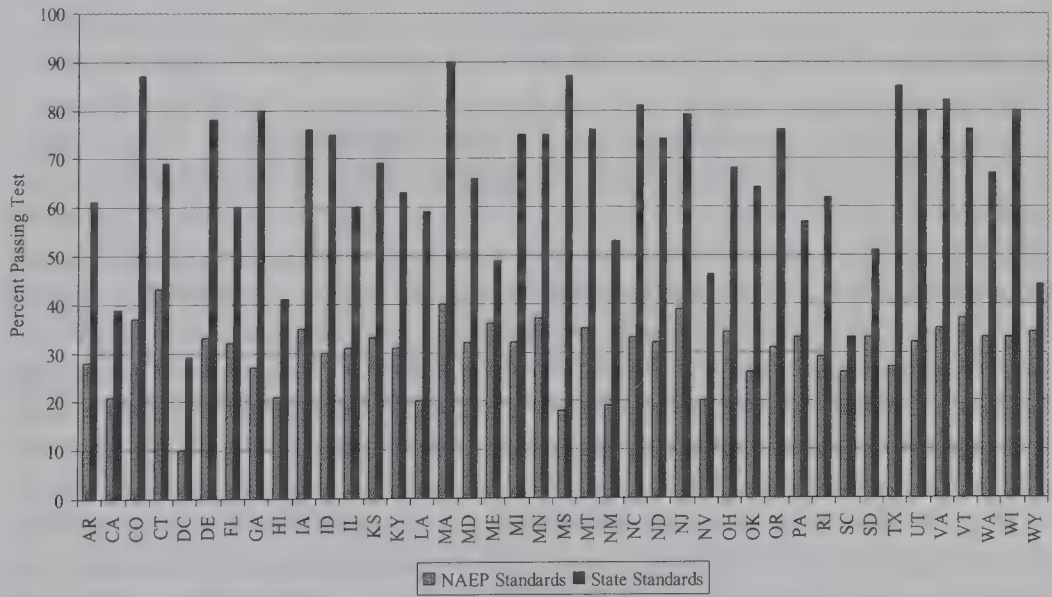


FIGURE 9 State and national proficiency score comparisons.

Country	1995	Country	2003
Singapore	590	Singapore	594
Japan	567	Hong Kong SAR <sup>1,2</sup>	575
Hong Kong SAR <sup>1,2</sup>	557	Japan	565
(Netherlands)	549	Netherlands <sup>1</sup>	540
(Hungary)	521	Latvia-LSS <sup>1</sup>	533
United States	518	England <sup>1</sup>	531
(Latvia-LSS) <sup>3</sup>	499	Hungary	529
(Australia)	495	United States <sup>1</sup>	518
Scotland	493	Cyprus	510
England	484	Australia <sup>1</sup>	499
Norway	476	New Zealand <sup>4</sup>	496
Cyprus	475	Scotland <sup>1</sup>	490
New Zealand <sup>4</sup>	469	Slovenia	479
(Slovenia)	462	Norway	451
Iran, Islamic Republic of	387	Iran, Islamic Republic of	389

■

 Average is higher than the U.S. average

○

 Average is not measurably different from the U.S. average

□

 Average is lower than the U.S. average

FIGURE 10 International comparison of average fourth-grade math scores (1995 & 2003).

SLIDING SIDEWAYS AND LOSING MOMENTUM TO JUDICIAL CONCERNS FOR EQUITY AND “ADEQUACY”

The nation’s most dramatic departure from convention in American school finance policy occurred during the 1960s and has continued for 50 years thereafter. This is the forceful emergence of the judiciary as a finance policy-setting agency. Initial post–World War II equal protection court cases concentrated on what a National Research Council report labeled “Equity I” (Ladd & Hansen, 1999), interdistrict, intrastate, difference in school district property wealth, and related differences in per-pupil revenue generating capacity. Second-generation cases, so-called Equity II cases, beginning in 1989 were thereafter filed in parallel with the interdistrict revenue capacity issues and continue to be filed to this day. Equity II cases, dealing with issues of adequate financing, invite the court to concentrate on a different question, not whether resources are equitably accessible to school districts but whether available resources are adequate to accomplish specified purposes. The latter issue demands a far more complicated level of proof than the former.



Equity I cases, concentrating on differences in wealth and spending between districts in a state, seemed to be justified morally and legally. These cases have generally been resolved favorably for the plaintiffs. Equity II cases, those dealing with adequacy issues, appear to have distracted state policy systems from important issues regarding school effectiveness. Adequacy cases are increasingly being decided in favor of defendants and plaintiffs eventually may be discouraged regarding the filing of such cases in the future.

### *The Successful Pursuit of Distributional Equity, Equity I*

Near the beginning of the 20th century, Ellwood Patterson Cubberley began to write scholarly pieces about the need to equalize the capacity of local school districts to generate revenue. Inequities were a function of unequal distributions of property wealth, and the more numerous and geographically small a state's school districts, the greater the probability that property wealth was ill-distributed among them. Cubberley and colleagues were persuasive, and legislators for 2 decades thereafter began to enact equalization provisions.

These plans, usually so-called Foundation Plans, were widely enacted during the first quarter of the 20th century. Their existence ensured that districts had equal access, at comparable property tax rates, to local property wealth, at least up to a per-pupil spending threshold that the state defined as a "Minimal Foundation." If a district's property tax base was insufficient to generate the state-specified foundation dollar amount, at the tax rate the state established, then the deficit was subvented to the district as a state financial subsidy.

Presumably a minimal foundation was the dollar level needed to ensure that students learned what was expected of them. However, the Minimal Foundation, more often than not, was a function of the status and solvency of the state treasury and was seldom an actual calculation of what was needed to educate a child. Still, Foundation Plans achieved a far greater degree of resource equality than the *ex ante* condition.

Even following state enactment of Foundation Plans, spending and resources inequalities remained. However, the Great Depression and World War II deflected the nation's attention away from schooling matters. Thus, it was in the post-World War II civil rights era that scholars rediscovered the inequalities that permeated school finance. Indeed, in the intervening quarter century encompassing fiscal duress and warfare, the disparities in local property wealth had been exacerbated, and now local school district spending differences were wider than ever.

In the 1960s, two scholarly groups became personally aware of the spending disparities, and separately, each constructed a legal theory to challenge the constitutionality of state school finance plans. Each theory was published in a prominent volume: Coons, Clune, and Sugarman (1970) authored *Private Wealth and Public Education*. Wise (1968) published *Rich Schools Poor Schools*. These two books,

written in isolation of one another, constructed a similar constitutional argument based on the "Equal Protection Clause" of the U.S. Constitution's 14th Amendment. These legal arguments became the basis for what emerged throughout the remainder of the 20th century as the largest judicial intervention in American education policy, short of *Brown v. Board of Education* and racial desegregation.

There were many ups and downs in the history of the equal protection legal movement. In 1983, the U.S. Supreme Court entered the fray with its decision in *Rodriguez v. San Antonio*, a narrowly decided decision that threatened to halt the entire finance reform movement by its negation of a Texas ruling favoring plaintiffs. However, proponents of education finance reform, thereafter, relied on state constitutional provisions to circumvent the federal precedent, and the equal protection suits persisted, often triumphed, in court. Among these decisions are famous landmarks for plaintiffs such as *Serrano v. Priest* in California, *Robinson v. Cahill* in New Jersey, *Seattle v. Washington*, and *Rose v. Kentucky*.

Equity I cases were, at least in retrospect, relatively simple. There were wealth disparities—often these were substantial disparities—among local school districts. Indeed, in Texas prior to subsequent litigation, the highest spending school district spent 25 times more money per pupil than the lowest spending low-wealth district. When suits were filed and trials were initiated, the evidence consisted of school finance experts explaining to the court that state foundation plans equalized only to a specific dollar level; thereafter, local wealth variations penetrated the arrangement, and herein was the crux of the disparity. In addition to expert testimony, it was possible to employ a wide range of statistical procedures to measure and display the degree of inequality. Berne and Stiefel (1983) authored what came to be the authoritative reference providing various definitions of equality and sophisticated statistical means for measuring degrees of inequality.

The Equity I remedy sought was often simple as well. The courts were asked to require the legislature to eliminate the inequity. That almost always required higher levels of state funding. No one really wanted to take money away from high-spending districts. Leveling up low-spending districts was, thus, the usual remedy. In addition, however, some states moved to restrict the ability of high-wealth districts to spend at their previous luxurious levels. In effect, their local control was compromised. The worst conflict of all, however, resulted when courts, or legislature, in search of a remedy required that some high-spending districts forego their expensive programs and forfeit some of their money to low-wealth districts. These recapture and redistribution decisions were legendary for the conflict they triggered in Texas.

By the mid-1990s equal protection trials had taken place in more than half the states. In most of these, plaintiffs prevailed. The cumulative result was a substantially greater level of interdistrict spending equality than had ever existed in the nation's history. Indeed, Murray, Evans, and Schwab (1988), after applying virtually all of the Berne and Stieffel equity measures and relying on sophisticated

econometric analyses, reported that the principal inequalities existed among, no longer within, states. To be a school child in Mississippi, Louisiana, Alabama, or New Mexico, all other things being equal, is not to have access to the same levels of school funding as in Connecticut, New Jersey, or New York. Indeed, New Jersey, the nation's highest per-pupil spending state, allocates twice what Mississippi is able to generate, a huge difference even when accounting for differences in living costs between the two locations.

Given the mobility of America's population, the need for national learning standards, significant interstate inequities are unjustified. The total costs to mitigate such disparities would not be great, and the regulatory and distribution mechanisms by which such could happen are easy to envision. Indeed, there may even be a constitutional interpretation that facilitates a federal government action to mitigate interstate revenue disparities (Goodwin, 2005).

There is one remaining spending inequality intradistrict inequality. In many large cities, senior teachers enjoy transfer privileges that they can exercise unilaterally. When senior teachers congregate in a select few schools, this condition can easily generate substantial per-pupil spending inequalities within a school district. These inequalities can exceed interdistrict disparities. However, here again, the technical means for eliminating these inequalities are simple. The principal obstruction is political.

### *Dysfunctional Judicial Efforts to Define and Attain "Adequacy"*

In the late 1980s and throughout the 1990s, state legislatures acted on federal government inducements and admonishments to adopt specific learning standards for students. Thus, for the first time in American history, standardized tests could be constructed and administered that actually measured that which a state specified was important for a public school student to be able to know and to do.

In part, state learning standards were operational. That is, they were intended as an integral component of an accountability system. If tests could be constructed to measure progress toward specific standards, then it became easier to hold school districts and schools responsible for children's learning. In part, however, the learning goals were aspirational. Legislators assuredly did not believe that every student would learn everything connected with mathematics, reading, science, and so forth. The learning standards were guides, not mandates.

Children's advocates, equal protection plaintiff attorneys, professional educators of all stripes, and social activists quickly seized on state learning standards as a means to leverage added revenues for schools. The logic was simple. If the state specified that Johnny and Suzy had to learn "X," then it only made sense, and constitutional sense at that, to ensure that Johnny and Suzy had sufficient

resources to learn “X.” However, whereas the logic is simple, the research challenges are daunting. No one knows what it takes for Johnny and Suzy to learn anything. Plaintiff attorneys attempted to demonstrate that low achievement, particularly among English Language Learning, handicapped, or minority students, was ipso facto evidence of districts possessing insufficient financial resources. However, judges came to understand that if student academic achievement was to be the definition of “adequate,” then there just might not be sufficient resources in the entire world to ensure that all children performed to a high standard on tests. Courts increasingly have been unwilling to define opportunity as test scores. Consequently, after enjoying early success in states such as Kentucky, New York, and Wyoming, courts have now begun to have second thoughts and adequacy, as an argument may have diminished attractiveness for plaintiffs (Lindseth, 2007).

The adequacy movement, Equity II, thus may be on its last legal legs. If so, its distractive capacity will be diminished. However, if it survives as a legal cause it could continue to have two kinds of deleterious effects. First, if legislatures see themselves as responsible for funding aspirations, then they will either render learning specifications vague, in which case accountability is eroded, or lower standards, in which case learning suffers. Either way, school effectiveness is damaged. This is not even to mention the millions of dollars that defendants routinely spend in defending the state against such lawsuits.

## AN UNBALANCED AND UNMEASURED NATIONAL EDUCATION REFORM PORTFOLIO

The United States has been actively engaged in education reform for a quarter of a century, since the 1983 publication of *A Nation at Risk*. However, the reform effort has been lacking on two crucial dimensions. First, the portfolio of reform strategies is badly slanted away from market force. Second, there is little by way of a systematic effort to appraise the consequence of reforms and thus little by way of an ability to profit from failure, to learn from successes, or to undertake midcourse corrections.

### *“A Nation at Risk.” The Mother of all Modern American Education Reform*

It is difficult to imagine that a slender government publication, *A Nation at Risk*, a document that was badly flawed analytically, could accomplish so much good in its wake. The central message in this Reagan-era document was that American school achievement had fallen so low as to jeopardize the nation’s economic well-being. The publication presented no persuasive evidence that American achievement had fallen from a prior point. Moreover, it provided no empirical



research regarding a direct link between national economic well-being and student academic achievement.

### *Elevated Expectations and Low Hanging Fruit*

Whatever its deficiencies, *A Nation at Risk* ignited a firestorm of education reform activity. State after state moved to upgrade high school graduation requirements, and colleges elevated admission standards. School districts eliminated electives and insisted on more rigorous programs of study for students to graduate. Testing was ramped up steeply. Textbooks were accused of having been dumbed down and thus were made more rigorous. Homework came back in favor. The school year was lengthened and the school day was extended. Physical Education was removed as a state-required subject. On and on the reforms proceeded. Virtually no low-hanging reform fruit was left unharvested.

### *"Alignment:" The Silver Bullet in the Privileged Portfolio*

*A Nation at Risk's* beneficial effects still can be felt. It is not difficult, for example, to trace the lineage of *No Child Left Behind* to *A Nation at Risk*. Still, by the late 1980s it was becoming evident that student academic achievement was not responding to quick fixes. A new theory of education reform was advanced and quickly adopted. The new explanation for lack of school productivity was the misalignment of instructional components (O'Day & Smith, 1993).

The article by Marshall Smith and Jennifer O'Day flashed around the education world every bit as fast as *A Nation at Risk*. To be sure it did not attract the massive attention of the popular media, and most members of the general public never knew of its existence. However, for professional educators, Smith and O'Day quickly assumed biblical significance. Their message was simple, but their effect was huge. What these two scholars suggested, while beguilingly simple, explained much. If students were not learning sufficiently, then perhaps schools were not instructing effectively. For schools to be effective instructional engines it altogether made sense that what teachers taught, what curriculum guides contained, what subjects schools offered, what standards administrators involved, what textbooks contained, what homework underscored, what state teacher certification required, and what schools of education promulgated for teachers should all be consistent and aligned with what state and district subject matter tests measured.

The O'Day and Smith hypothesis led to a flurry of federal, state, and local school district activity. New textbooks were commissioned, new curriculum guides were written, new teacher training requirements were created, and ever more sophisticated means were constructed for measuring the degree to which alignment existed (Porter, 2002). Alignment was also attractive to those who desired accountability because it made clear that there were standards and objective measures against

which the effectiveness of a school could be judged. Of course, there was nothing in the alignment strategy that specified consequences for school where everything was aligned, but students still did not learn. Had such consequences been included, it is unlikely that alignment would have been the attractive reform magnet that it was.

The Clinton administration climbed aboard the alignment bandwagon. Marshall Smith became the deputy secretary of the Department of Education. Legislation was enacted establishing a National Goals Panel to which states submitted their learning objectives for federal approval. If states could just get the goals correct, and if there were sufficient alignment of all the instructional components with those goals, then, voilà, student achievement would be elevated. The Elementary and Secondary Education Act was amended to encourage schools receiving aid for low-income students to adopt "Whole School Programs." Added federal legislation was enacted providing school districts with Comprehensive School Reform funding. The notion here was that it took a village, at least a village inside of a school, to ensure that students learned (Clinton, 1996). Of course, all the parts of the village had to act in consort. All the parts had to be aligned.

### *Slighted Strategies*

There was much about alignment that was, and still is, attractive. It is a most logical way of viewing the operational and instructional world of schooling. However, alignment as a panacea is deficient. It assumes that schools are filled with adults who, if schooling components were simply all lined up correctly, would be eager to teach and eager to learn how to teach more effectively. Alignment had little to say about motivation. It is this oversight that attest to the unbalanced nature of the reform portfolio. Where are performance incentives? Where are the consequences for consistent poor performance?

*Markets and incentives.* Under current conditions in American public education, if students do not perform well or if parents are dissatisfied, there are limited options. Few teachers or administrators lose their jobs because of underperforming or under-chosen public schools. Thus there is a question as to whether greater amounts of competition might act as an incentive for educators to strive more to gain elevated achievement and parent satisfaction. After all, if a private school is not chosen by households, it runs the risk of going out of business and its employees run the risk of losing their jobs.

To some degree, the United States has experimented with a greater amount of privatization in public schools. This is the charter school movement, and it now appears that about 5% of America's public school stock is charter schools. However, this is hardly sufficient competition to test whether parents can shape a

new system by voting with their feet. The number of voucher plans in the play in the United States is smaller by far, and certainly is not now capable of injecting significant competition into the system.

Any balanced reform portfolio would allocate substantially greater resources and effort toward the development and appraisal of a competitive sector to determine if competition had a chance of elevating student performance and parental satisfaction.

Performance incentives also represent an insufficiently tried instrument for possibly enhancing the effectiveness of schools. States have the led the way, notably Florida and Texas, in designing and attempting to evaluate performance incentives for teachers. The federal government supports several experiments on performance pay, and that is to be applauded. Congress enacted the Teacher Incentive Fund, authorizing 34 pay-for-performance projects in school districts and charter schools throughout the United States. However, this federal effort is unusually politicized, has not been undertaken systematically, and is already known to have triggered several thoughtless and ill-conceived operations that are more likely to give performance pay a bad name rather than advance the game.

*Research and development.* Even if the nation's education reform portfolio were rendered more complete by the inclusion of market competition and incentives, it would be unlikely that the policy community would learn a great deal. Systematic efforts to appraise outcomes and undertake rigorous empirical inquiry regarding education reform are badly under funded. A bare-bones research and development (R&D) effort would utilize 1% of operating funds to conduct research. In American education this would translate to \$3.3 billion annually. Including every possible combination of federal and philanthropic foundation research funding maximally generates one tenth of that amount. If funding for the NAEP is removed from this calculation, then R&D funding is about \$200 million, or approximately .006 of operating costs.

## ABSENCE OF A NATIONAL VISION AND A NATIONAL STRATEGY

The helter-skelter higgledy-piggledy nature of U.S. education finance and U. S. education policy is generally accepted. It is frequently described, often lamented, and just as frequently accepted as a *fait d' accompli*.

### *Centrifugal Forces: Decentralization and Technical Uncertainty*

The crazy quilt policy landscape is conventionally attributed to the vastly decentralized nature of our education policy-making machinery. Federal, state,

and local authorities all have much to say in the shaping of schools, and special interest groups operate across all levels with powerful means for insinuating their often self-serving agendas into the policy-making broth. The booming buzzing confusion is made all the more confusing by the absence of empirical research findings that could compose a technical base for which to construct productive education policy. Hence, it is understandable that courts wander into impenetrable evidentiary thickets related to adequacy, and policymakers grasp at brass rings and silver bullets such as alignment, whole school reform, scripted instruction, scientific management, reading recovery, phrenology, or any one of hundreds of other short-lived or low-performing fads. There appears to be little by way of a beacon providing guidance on the proper direction for the nation when it comes to education. Conversely, the initial impression is that education policy is a function of a giant centrifuge forcefully propelling all to the periphery where it is subject to little coordination and accountability.

### *Centripetal Forces: What to Do When You Do Not Know What to Do*

However, all is not bleak. There are, few to be sure but nevertheless positive, instances of national movements that have born productive practical and policy fruit. For example, the NAEP has existed for 40 years, and without it the nation would have few if any means for measuring academic progress or the lack thereof. There are national examinations applicable to college admission, the SAT and ACT. Nationally distributed textbooks contribute to a greater commonality in the school curriculum than is frequently acknowledged. Models for teacher training and licensing, rightly or wrongly, are generally common across states. There is a National Board of Professional Teacher Standards. Education finance mechanisms display remarkable commonality. The point here is not to argue the merits of any particular common dimension or national element but, rather, to emphasize that there are centripetal forces than can be harnessed in pursuit of significant policy objectives.

For reasons explained in the following section of this article, it is highly unlikely that the political system, at any government level, can overcome existing structural obstacles and procedural roadblocks, and just raw institutional inertia, to formulate an effective and comprehensive education policy. Regardless of whatever political party dominates national or state governments, a national education strategy will simply not emerge politically. A national disaster such as extremes of global warming or expanded warfare might alter the scenario and cause the United States to think nationally. However, failing such cataclysmic conditions, the crazy quilt and incoherent present pattern will prevail, and the best one can expect is narrow increments of change from time to time.



From time to time, however, national changes do occur in American education. One need only reflect upon the dynamics that gave birth to *A Nation at Risk* to realize that, under selected circumstances, the public can be galvanized to endorse substantial directional change. Thus, assuming public support, what should comprise the agenda?

There are three activities worthy of consideration in a national education action agenda: overcoming the “Adequacy Distraction,” expanding the reform portfolio, and instilling a national mindset of experimentation and continuous learning.

In a paper recently commissioned by the Gates Foundation—sponsored School Finance Study Group, Guthrie and Hill specify a set of steps that could be useful in converting the education system into a mechanism for overcoming technical uncertainty, a mechanism that leads to a continuous cycle of experimentation and improvement. The Guthrie and Hill paradigm for acting when there are no clear directional signals includes activities such as following:

- Placing resources close to students and in organizations that can be held accountable.
- Rendering resource distributions transparent.
- Encouraging widespread experimentation with competing improvement models.
- Establishing financial accounting and performance related databases that facilitate productive program evaluation and research.
- Installing performance rewards as incentives for schools and professional educators.
- Planning intentional experiments on important instructional and structural issues.
- Constructing accountability consequences that concern employees, not simply clients.

### THE UNLIKELY PROSPECT OF NEAR TERM POLITICAL CONSENSUS

America’s political system will not quickly face the challenge of linking school resources to elevated school achievement. The prospect of achieving substantial political agreement on the mission is small. The issues are remarkably fractious, there is little empirical evidence to act as a policy guide, there are daunting structural impediments to public engagement with the problems of education, and unless it is a matter of “high politics” involving the most influential levels of government and elected officials, education special interests intractably dominate the education policy political landscape. *If any progress is to be made, the responsibility will fall to externally organized actors.*

Here are the barriers to near term political consensus.

*The value quagmire.* While the education system begs for attention to matters of productivity (e.g., substituting capital for labor, determining the effectiveness of performance incentives, or experimenting with market competition), much of the policy system is locked into debates regarding values. As evidence, witness the continuing debate about evolution and creationism in Kansas and other states. Take, as another example, whether a New England middle school should issue condoms to middle school adolescents. Remember the furor in New York City schools over whether an adopted textbook should make reference to same-sex partners?

*The absence of evidence.* But even if education politics were not ensnared in society's unresolved value conflicts, there is still precious little reliable research evidence to guide policymakers. Only the Tennessee Star study can be said to provide experimental evidence. Most of the remainder of what passes for research in evidence does not pass methodological muster.

*Structural crazy quilt.* The structure of American education governance, coupled with a variety of procedural dynamics, such as Progressive Era efforts at depoliticization (e.g., separation from municipal government, nonpartisan school boards, and off-year elections) and modern era collective bargaining, renders it difficult to achieve a citizen consensus regarding education policy. The U.S. Constitution is silent regarding education, and that condition coupled with the 10th Amendment's empowerment of states, devolves education authority to state governments. States, historically, have depended on local school districts as their operational agencies. The result of such complexity is that the United States has almost 14,000 local districts, most with elected or appointed local school boards; 50 states with a variety of governance mechanisms; and a federal government whose potential influence is substantial but whose actual authority is cumbersome and inconsistent.

*Microdecoupling.* This crazy quilt pattern of governance and operational complexity contributes to conflict and privileges the status quo. The principal problem is the misalignment between those who bear the burden of financial support and those who receive the benefits of a current or anticipated governmental arrangement. When it comes to education policy, transaction costs for citizen political engagement in issues is unusually high, the payoff is uncertain and remote, and school district employee returns to their engagement are unusually high. Take a local property tax increase as an example. While holding the prospect of raising

millions of dollars in aggregate revenue, a tax increase imposes a burden of but a few additional dollars on each individual household, a burden perhaps not justifying owner resistance given the information costs and likely amount of effort involved in opposition. Conversely, for a teacher union and its individual members, working hard politically for a tax increase makes enormous good sense, given the likely personal rewards involved. This condition is known as microdecoupling in political economics (Wolfe, 1997).

In short, unless an issue becomes one of high politics, involving the president and congressional leadership, becomes a component of a national political party platform, or an issue adopted by one or more governors and high-level state legislative leaders, then the politics of education at all levels—federal, state, and local—are dominated by special interest groups. Few education employees, or at least few teacher leaders, are eager to have the rigorous measurement and accountability that ultimately will be needed to render America's schools effective. Education does not heal itself. Indeed, it will hardly even diagnose itself. Hence, in the next 5 or 10 years, if there is to be any progress whatsoever regarding education productivity, it will more likely come from external pressures brought to bear on high levels of the political system. Constructing a National Strategy

Figure 11 summarizes factors related to eight of the nation's most prominent 20th-century education reform efforts. These eight have had dramatic impacts, either at the time of their initiation or still.

### *Retrospective: Learning From Case Examples*

Among the conditions that can be deduced from the foregoing display is that only seldom do educators initiate significant reforms. Generally, reforms stem from external societal pressures. The combined reform participation of business officials and academics can be particularly influential. Philanthropic foundation resources can facilitate change substantially, if aimed in a productive direction. Media involvement is important, and, finally, credible evidence can assist.

Here is a specific example in which many of the aforementioned reform facets came together.

In the 1970s an intellectual foundation had been constructed by the previously mentioned academic writings of John E. Coons, William H. Clune, and Steven D. Sugarman and of Arthur Wise to provide a constitutional basis for challenging interdistrict wealth disparities. James A. Kelly, then a Ford Foundation program officer, took up the cause and initiated a philanthropically financed set of activities that saw the challenge of school district resource inequality through to a most successful conclusion. These are the eight principal components in the Ford Foundation financed school resource equality strategy:

- **Organized for, But Not of or by Government:** The school finance equity campaign was always clear that it was intended to influence government, be it a court or a legislature, but that it was not going to accept government support or resources.
- **Philanthropic, Business, and Academic Alliances:** The finance equity campaign maintained a large, focused, nonpartisan tent. It accepted no issues, others than those directly on its charter. However, if a business, a foundation, or a group of academics was aligned with its purposes, then it was interested in their participation.
- **Assembling Champions:** A conscious effort was made to identify and recruit to the cause highly visible attorneys, academics, business officials, and legislative and executive branch leaders who were willing to make a sustained commitment of their time and expend an amount of their political capital on behalf of the campaign.
- **Constructing Informational and Professional Networks:** Newsletters, the conscious circulation of relevant publications, commissioning of research papers, and the subsidization of professional meetings were all undertaken. A cadre of consultants was enlisted and was on call to attorneys or others throughout the nation who asked for assistance.
- **Recruitment and Training of Attorneys, Scholars, and Reporters:** A sustained effort was made continually to identify additional talent and expand the networks of informed and able attorneys and school finance analysts who could file cases, conduct equity studies, engage in legislative briefings, and design legislation. Workshops were organized for members of the media to provide them with background regarding the issues.
- **Constructing Legislative Models and Judicial Portfolios:** Sample school finance reform bills were drafted and templates and briefs were written by nationally expert attorneys to be used as models by lawyers throughout the states who were otherwise insufficiently informed regarding the larger legal issues.
- **Policy Research, Public Information, and Lobbying Missions:** A constant flow of information memos regarding the cause in general and specific states, in particular were continually being prepared and made available to those engaged in lobbying in state capitals.
- **National and Regional Conferences:** The Ford Foundation routinely organized and financed conferences, some national, some regional to convene participants. These were in substantial measure for information and networking purposes. They also served as a motivational device, cheerleading for those in the front lines of litigation legislation. Each favorable trial decision, each enacted bill served as a justification for celebration.
- **Media, Media, and More Media:** Throughout the aforementioned activities and events, there was a never-ending media campaign. Articles, op-ed pieces, human interest stories, trial snippets, legislative status reports, factoids, anecdotes,



FAQs, advisories, and TV and radio announcements at the state and national level were continually in preparation by a nationally experienced publicity corporation paid for, indirectly, from Ford Foundation funding.

- Sustained Commitment: One of the components of the equity campaign's success was the knowledge that it was a long-term effort. Knowing that an individual or an organization would have financial support, not for a year or 2 years but for 3 to 5 years, provided a level of security that facilitated recruitment of able individuals to the cause.

### *Prospective; Framing and Implementing a National Agenda for the Future*

Dare one think of a national effort to enhance education policy, to move the nation closer to a coherent set of strategies that hold the prospect of significantly elevating academic achievement? If anyone, or any group, is sufficiently bold or naive to think such is doable, then there follows a few lessons from prior efforts that might apply to the future.

## REFERENCES

- Berne, R., & Stiefel, L. (1983). *The measurement of equity in school finance: Conceptual, methodological, and empirical dimensions*. Baltimore: Johns Hopkins University Press.
- Clinton, H. (1996). *It takes a village*. New York: Simon & Schuster.
- Coon, J. E., Clune, W. H. III, & Sugarman, S. D. (1970). *Private wealth and public education*. Cambridge, MA: Harvard University/Belknap.
- Costrell, R. M., & Podgursky, M. (2008). Peaks, cliffs & valleys: The peculiar incentives of teacher pensions. *Education Next*, 8(1), 22–29.
- Goodwin, L. (2005). *Education, equality, and national citizenship* (Paper No. 30). Center for the Study of Law and Society Jurisprudence and Social Policy Program, University of California, Berkeley.
- Guthrie, J. W., & Paul, T. H. (2007). *Making resource decisions Amid Technical University*. Commissioned for the School Finance Redesign Project Center on Reinventing Public Education, University of Washington.
- Ladd, H. F., & Hansen, J. S. (1999). *Making money matter: Financing America's schools*. Washington, DC: National Research Council Press.
- Lindseth, A. (2007, September 17). A reversal of fortunes: Why the courts have cooled to adequacy lawsuits. *Education Week*, 27(3), 32.
- Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children*, 5(2), 113–127.
- Murray, S. E., Evans, W. N., & Schwab, R. (1988). Education finance reform and the distribution of education resources. *American Economic Review*, 88(4), 789–812.
- O'Day, J. A., & Smith, M. S. (1993). Systemic reform and educational opportunity. In S. Fuhrman (Ed.), *Designing coherent policy: Improving the system* (pp. 250–312). San Francisco: Jossey-Bass.

- Porter, A. C., & Smithson, J. J. (2002, April). *Alignment of assessments, standards and instruction using curriculum indicator data*. Paper presented at the AERA meeting, New Orleans.
- Porter, A. C. (2007, Fall). Rethinking the achievement gap. *PennGSE: A Review of Research*, 5(1), 1.
- Rothstein, R., & Miles, K. (1995). *Where's the money gone? Changes in the level and composition of education spending*. Washington, DC: Economic Policy Institute.
- Wise, A. E. (1968). *Rich schools: Poor schools*. Chicago: University of Chicago Press.
- Wolfe, C., Jr. (1997). *Markets or governments: Choosing between imperfect alternatives* (2nd ed., pp. 2, 43). Cambridge, MA: MIT Press.

# A Framework for Choice Remedy Litigation

Clint Bolick

*The Goldwater Institute Phoenix, Arizona*

Although school choice proponents have generally been on the offensive in legislative arenas over the past 2 decades, they have played almost constant defense in the judiciary, seeking to prevent courts from undoing school choice programs. Opponents typically wield state constitutional provisions against school choice programs. Properly construed, such provisions often are intended not to thwart but to secure educational opportunities. School choice supporters should consider taking the offensive, applying such provisions toward their intended ends by challenging defective schools and seeking meaningful remedies for children trapped in them. Choice remedy litigation can provide an effective complement to legislative efforts in the larger campaign to secure for disadvantaged children the precious educational opportunities that are their constitutional right.

Ever since the first urban private school choice program was enacted nearly 2 decades ago, legal challenges have been a constant feature of the terrain. Parental choice advocates have successfully fended off First Amendment challenges, culminating in *Zelman v. Simmons-Harris* (2002)<sup>1</sup> but have met with less success thus far in defending programs against state constitutional challenges.

It is odd in a nation doctrinally committed to equal educational opportunities (and most of whose state constitutions expressly provide a right to education) that advocates of expanded choices should find themselves constantly on the legal defensive. Given that appalling educational inequalities continue to prevent us from fulfilling this sacred moral promise to our nation's children and that courts exist to uphold fundamental rights and to dispense justice and equity, advocates of parental choice should not consider it a natural condition to be on the defensive in

---

Correspondence should be sent to Clint Bolick, Goldwater Institute, Scharf-Norton Center for Constitutional Litigation, 500 East Coronado Rd., Phoenix, AZ 85004. E-mail: [cbolick@goldwaterinstitute.org](mailto:cbolick@goldwaterinstitute.org)

<sup>1</sup>I recount the successful initial 12-year litigation effort to defend school choice programs in Bolick (2003).

the legal arena. Most to the point, we should not permit constitutional guarantees of educational opportunity to be used to thwart such opportunity. Yet we allow that to happen when we cede the legal arena to our foes.

Opportunities abound for advocates of parental choice to advance their cause through litigation. In this article, I focus on the most promising approach for systemic change: choice-remedy litigation using state constitutional guarantees and building on funding equity jurisprudence.

For more than 35 years, courts across the nation have applied state constitutional guarantees regarding education to increase funding for public schools. In some instances, those who favor greater parental choice have attempted to influence the course of such cases, sometimes by opposing them and sometimes by seeking to intervene to advocate a different remedy. Mostly they have sat on the sidelines, allowing the groups who are prosecuting such lawsuits to define the terms of the debate in terms of money rather than meaningful educational opportunities. Unfortunately, the massive increases in funding that have resulted from such lawsuits rarely have trickled down to the intended beneficiaries of the educational guarantees.

That will remain the case until advocates of parental choice enter the fray in a serious and systematic way. This article is intended to sketch a path for parental choice advocates to effectively invoke educational guarantees to increase educational opportunities for the children who most need them.

## FROM EQUITY TO ADEQUACY TO CHOICE

The earliest school finance equity case was filed in federal court. The U.S. Supreme Court rejected the notion of an affirmative right to education in the U.S. Constitution in 1973 (*San Antonio v. Rodriguez*). Under that precedent, to satisfy the dictates of equal protection under the 14th Amendment, a state need only demonstrate a “rational basis” for the classifications it creates in the education context—a standard so deferential to that in reality it does not require government decision makers to articulate a basis for its classifications at all, much less one that is in any sense truly rational.

Since that decision, advocates of school finance equity have focused on state courts and constitutions to achieve their objectives. The school finance equity campaign has been one of the most successful of the efforts by liberals over the past 40 years to advance their ends through state courts, rather than through a federal judiciary that has turned increasingly conservative.<sup>2</sup>

---

<sup>2</sup>Only recently have conservatives and libertarians begun systematically to focus on state constitutions to advance freedom. The Goldwater Institute was the first market-oriented policy group to



The first two successful school finance equity cases took place in the early 1970s in New Jersey (*Robinson v. Cahill*, 1973) and California (*Serrano v. Priest*, 1971). Like many state constitutions, New Jersey's contains an explicit education guarantee, specifically entitling all children to a "thorough and efficient" education. By contrast, California's constitution did not contain an express education guarantee. But as part of our system of federalism, states are free to interpret their own constitutions to confer greater protections than the federal constitution, even where the language in the two constitutions is exactly the same. The California Supreme Court did so, recognizing education as a "fundamental" constitutional right. Under that standard, government classifications can survive judicial scrutiny only if they are narrowly tailored to a compelling governmental interest. Applying their state constitutions, the New Jersey and California Supreme Courts invalidated their respective state finance systems.

The first lesson that parental choice advocates can learn from the finance equity cases is that judicial action can bypass, compel, or at least hasten legislative action. Not all were successful: Several state courts ruled the question of funding equity "nonjusticiable," holding that no matter how explicit the education guarantee, the state constitution vested the matter entirely to legislative discretion. But enough of the lawsuits were successful to effectuate a fundamental change in education finance across the nation, largely accomplishing the movement's three signal objectives: (a) the displacement of property tax-based school financing with financing from state sources, (b) the displacement of primarily local responsibility for school financing with primarily state responsibility (along, of course, with greater control), and (c) dramatically increased funding, particularly for property-poor school districts. Left only to the legislative arena, finance equity advocates might never have accomplished all of those changes, or at least not in so short a period, given the powerful forces arrayed in support of the status quo. But judicial action forced recalcitrant legislatures to act and created an inexorable national tide of education finance reform.

The finance equity advocates deployed three important weapons that were crucial to their success. First was a cadre of tenacious, committed, skilled lawyers who relentlessly litigate finance equity cases and who in turn developed a core of "experts" available to testify in cases across the country. Second was an aggressive campaign in the court of public opinion. Third was the "sweetheart" lawsuit—cases in which government defendants were all too happy with a finding of constitutional deficiencies that would reap them millions or billions of additional taxpayer dollars. Parental choice advocates should be able to acquire the first two weapons but rarely if ever the third. Even if they can find states with sympathetic

---

launch a litigation program, the Scharf-Norton Center for Constitutional Jurisprudence, to focus almost exclusively on vindicating freedom protections in the state constitution (see Bolick, 2007).

attorneys general, they can count on powerful interests (such as teacher unions and school boards) to intervene as defendants and mount a vigorous defense.

Typically, the finance equity cases proceeded by showing large funding disparities between property-rich and property-poor districts and seeking injunctive relief. In theory, the injunctions left discretion in the hands of the legislature, but in reality they were a loaded gun: Solve the problem, or else. Legislatures eventually complied, raising taxes and pouring massive new funding into property-poor school districts. Per-pupil funding in such districts has increased dramatically (in some instances to \$20,000 per student). Meanwhile, in some states—most infamously, New Jersey—courts for many years have maintained jurisdiction over school funding, even to the level of minutiae. Hence, even with a number of court losses, finance equity advocates have succeeded beyond their wildest dreams.

But what the finance equity advocates have not been able to deliver—if it ever was their intended goal—is genuinely improved educational opportunities for disadvantaged schoolchildren (Hanushek, 2006). Over time, massively increased funding reaps diminishing returns, with school bureaucracies, personnel, and vendors (not to mention the lawyers) displacing needy schoolchildren as the true beneficiaries of the public largesse.

As funding inequities began to disappear—indeed, in many states, state funding for urban school districts significantly exceeds median district funding—advocates of yet greater public funding altered their legal theories to fit changed circumstances. The focus on funding equity began to shift to educational “adequacy” (see, e.g., Heise, 1995). Now the proof centered not on funding disparities but on the failure of students, regardless of how much money was being spent, to succeed academically. But the remedy remained the same: more money for “overburdened” schools.

Again, the plaintiffs succeeded in enough cases to get the spigot running again. Over the past several years, advocates of increased funding have prevailed in New York, Texas, and other states. And again, increased funding has not been accompanied by commensurate improvements in system accountability or student achievement.

That failure, especially in states that have traveled furthest down the road of increased funding, would seem to open the door to parental choice remedies. The equity and adequacy lawsuits are seriously flawed in multiple respects. First, the intended beneficiaries of the state constitutions’ education guarantees are not school districts but *children*. But children thus far have been mere props in the quest to secure ever-greater funding for school systems. Second, and related to the first, school districts are not *victims* of constitutional malfeasance but perpetrators of it. They are, at the very least, the state’s agents in delivering on the constitutional obligation to provide educational opportunities. Yet they show up in equity and adequacy lawsuits not as defendants but as plaintiffs. Third, the remedy defies

the most basic requirement of equity because it is grossly mismatched to the constitutional violation: Instead of providing immediate, make-whole relief to the victims, it showers dollars upon constitutional tort-feasors.

Unfortunately, advocates of greater funding have so dominated the legal arena and the terms of the debate for so long that they have turned the ordinary rules of equity upside down in *Alice in Wonderland* fashion: What in any other area of the law would be unthinkable now is commonplace; what should be commonplace is deemed radical.

To put the situation into perspective, I like to use a simple analogy from the context of product liability—which really is what we’re dealing with here. Let’s say a consumer purchases a car and receives from the manufacturer a warranty of “thorough and efficient” transportation. It turns out that the car is a complete lemon. The manufacturer attempts to repair it to no avail, leaving the consumer with no transportation at all, much less something thorough and efficient.

If the consumer went to court, what would a court do to redress the violation? What a court emphatically would not do is to award billions of dollars to the automobile manufacturer in the hopes that in this decade or the next it might produce a thorough and efficient automobile that it might provide to the consumer. Rather, it would give the purchaser her money back, which she can use at once to select a better product. The question is not a close one. Yet in the topsy-turvy world of school litigation, the first remedy is ubiquitous, whereas the second is dismissed as—gulp—judicial activism.

In reality, a “choice” remedy is not unknown even in education. Under the federal Individuals with Disabilities Education Act (IDEA), all disabled children are guaranteed a “free appropriate education.” In the first instance, public schools have the obligation and opportunity to provide an appropriate learning environment. But if they fail to do so, the U.S. Supreme Court has ruled unanimously that they must provide it at public expense in a private school chosen by the parents (*Florence County Sch. Dist. No. Four v. Carter by and through Carter*, 1993). Indeed, the more than 100,000 disabled children attending private schools under this interpretation of IDEA compose the nation’s largest parental choice program.

Parental choice advocates should endeavor to convince state court judges that they should interpret their own constitutions to provide precisely such immediate and meaningful relief. Indeed, even in states where funding equity or adequacy decrees are in place, parental choice advocates can argue that choice is an essential *interim* remedy; while the legislature complies with court orders and greater funding and accompanying reforms work their presumed magic, students should not be forced to remain in schools that are demonstrably inadequate. Parental choice advocates can show that even a temporary deprivation of educational opportunities can constitute irreparable injury and moreover can demonstrate, drawing upon experiences in Milwaukee, Florida, and elsewhere, that parental choice drives systemic accountability and reform.

Two such efforts along those lines were prosecuted in the early 1990s by the Institute for Justice—one in Chicago and the other in Los Angeles. Both failed in court (*Jenkins v. Leininger*, 1995).<sup>3</sup> In Illinois, the state constitutional guarantee of a “high-quality” public education was deemed aspirational only and therefore nonjusticiable. In California, the constitution was interpreted to preclude the voucher remedy. Despite the adverse court rulings, the cases were enormously successful in the court of public opinion, reaping a favorable headline in *USA Today*, editorial support from the *Washington Post*, and prominent coverage by national television media. The terms of the public debate over parental choice began to shift, linking the interests of disadvantaged inner-city schoolchildren with greater school choice. In turn, where only one urban school choice program (Milwaukee) existed prior to the lawsuits, several states and Congress enacted more than one dozen programs across the nation in the following decade.

Still, the last thing the parental choice movement needs is to invest precious resources in quixotic lawsuits. One of the frustrating but important realities we need to confront is that even as the appeal of school choice increasingly transcends class and philosophical boundaries, for many in positions of power the issue remains fiercely partisan and ideological. Thus, perversely, many of the same judges who are quick to recognize a central and activist role for the judiciary in enforcing state constitutional education guarantees often are ideologically opposed to parental choice. Likewise, judges who philosophically inclined toward parental choice tend to be deferential toward legislative prerogatives. The success of choice remedies depends on intellectually honest judges who are willing to vigorously yet objectively enforce constitutional guarantees.

What if anything has changed since the early 1990s to justify a renewed investment in litigation as a major part of the parental choice arsenal? At least five things.

First, conditions continue to deteriorate in inner-city public schools, with little to show for massive increases in public funding. Things had to get worse before they could get better—and they have. Many who genuinely believe in equal opportunity are growing more open to parental choice.

Second, advocates for increased public funding have unwittingly opened the legal door to choice remedies. The shift from equity to adequacy has created a favorable legal terrain for parental choice advocates, for a choice remedy fits much more naturally (as a permanent, partial, or interim remedy) than increased funding to districts that fail to meet constitutional standards.

Third, the progress of the movement toward educational accountability, abetted by the accountability requirements of the No Child Left Behind Act (NCLB),<sup>4</sup>

---

<sup>3</sup>The California decision is unpublished.

<sup>4</sup>In addition to the accountability requirements that are helpful to choice advocates in identifying failing schools, NCLB presently includes a guarantee of public school choice for children who are



has fueled the development of state standards for academic performance. Courts understandably often are reluctant to create standards by which to measure whether the state's constitutional obligations are being fulfilled. Now, with states setting their own standards for educational adequacy, parental choice advocates simply can apply those standards—which serve as the proverbial “smoking gun”—to establish the state's liability in failing to provide a constitutionally adequate education. That leaves to parental choice advocates the principal task of demonstrating that choice is the proper remedy.

Fourth, school choice now is a proven solution to the ills of inner-city public education. We can deploy our own cadre of experts to demonstrate that choice is the *only* remedy that *immediately* allows children to leave failing schools and enter better performing schools and that choice instills accountability and provides a catalyst for improvement in the public school system.

Fifth, after several years of legislative successes, opponents are striking back. This year and next may witness, for the first time, a net *decline* in the number of private school choice programs and the children able to utilize them, as a result of court challenges, voter initiatives and referenda, and shifting legislative majorities. In Utah, for instance, opponents successfully referred to the ballot the nation's first universal school choice program and scored a resounding 62–38 percentage victory at the polls. A carefully developed litigation program, combined with an aggressive campaign in the court of public opinion, is essential to preserve and accelerate the momentum of the school choice movement and the precious opportunities it is poised to deliver.

## LITIGATION LOGISTICS

Advocates in nearly all states should consider choice-remedy litigation. Obviously, the states that could benefit most are those with serious education problems and few prospects for achieving school choice through normal political processes. But states without troubled urban school districts can consider such lawsuits on a smaller scale, and states with existing limited school choice programs may enjoy an advantage in the litigation arena if the positive effects of choice are well known.

---

enrolled in schools that fail to make adequate yearly progress for two consecutive years. Few among the many eligible children have availed themselves of such options for a variety of reasons, including the failure of school districts to publicize the options (as the law requires them to do) and the lack of adequate school alternatives. Unfortunately, NCLB does not provide a private right of action to enforce the choice options and, of course, does not include private schools as options. The Alliance for School Choice currently has a complaint pending before U.S. secretary of education Margaret Spellings asking her to enforce the public school choice options for California children in the Los Angeles and Compton school districts.

The experiences of the funding equity and adequacy cases as well as the early voucher-remedy cases are instructive in guiding future efforts to achieve choice remedies under state constitutional education guarantees. Two absolute prerequisites exist before advocates should seriously consider filing a choice-remedy lawsuit in a given state: an enforceable education guarantee, and the availability of a choice remedy under the state constitution.

With regard to the first prerequisite, it is most useful to have a clear articulated guarantee (particularly one that is normative, such as “thorough and efficient” or “high quality”) that the highest court in the state has found to be justiciable. But it is enough, to at least consider going forward, that some guarantee exists and that the courts have not ruled that the clause is *not* justiciable. In most states, the equity advocates or others have resolved those questions one way or the other.

The terrain is less certain with regard to the permissibility of a voucher remedy. Only two states—Michigan and Massachusetts—have state constitutions that clearly preclude publicly funded private school choice altogether. Two others—Wisconsin and Ohio—have upheld school vouchers. The other states fall somewhere in-between.<sup>5</sup> The most common obstacles are the so-called Blaine Amendments, which are found in most state constitutions and prohibit aid to sectarian schools. Blaine Amendments should not necessarily deter school choice advocates, both because they can be construed to permit aid to students (as in Wisconsin and Arizona). To the extent they are applied to discriminate against religious school choices, they may violate the nondiscrimination guarantee of the First Amendment – an issue that school choice advocates are anxious to bring to the U.S. Supreme Court. Even in states that clearly or apparently prohibit private school choice, the effort may be worth pursuing in order to support change of the constitutional rule or its interpretation; or to set up a Blaine Amendment challenge in the U.S. Supreme Court.<sup>6</sup>

Once the basic legal parameters are established, the choice advocates should determine the factual predicate to establish a constitutional violation. Increasingly, especially in accord with NCLB, states have established accountability systems that assign grades to schools. Ideally, the system will be one like Florida’s, which ranks schools using grades from A to F, or New Jersey’s, in which the state legislature has given definition to the constitution’s education guarantee through proficiency tests, the results of which are available school by school. NCLB rankings, which measure “adequate yearly progress,” are not necessarily a surrogate for successful or failing schools (though schools that have failed to make adequate yearly progress for several years in a row safely can be said to be failing schools). But absent state standards that can be used to determine the identity of failing

---

<sup>5</sup>For a state-by-state assessment of the constitutionality of school choice, see <http://ij.org/pdf/folder/schoolchoice/50statereport/50stateSCreport.pdf> (Institute for Justice Web site).

<sup>6</sup>For a more extensive discussion of Blaine amendments, see Bolick, (2003).

schools, prospective litigants might consider using National Assessment of Educational Progress scores. If the state itself does not classify schools as “failing,” choice advocates will have to work carefully with experts to determine a defensible standard for identifying failing schools.

Using the state’s own school performance data is an excellent way to establish liability, because the data provide an objective measure created by the defendants themselves. Another possibility exists in states in which successful adequacy lawsuits have been litigated: Plaintiffs seeking choice remedies can build on already-existing findings of liability and argue that existing remedies are inadequate.

The most likely choice remedy litigation will take the form of a direct lawsuit. However, in states with existing lawsuits, choice advocates might consider intervening in those lawsuits to seek a different or interim remedy or, if the existing lawsuit is a class action, seeking to remove families from the existing class. The rules for intervention vary by state; generally, all require prompt action to intervene. But even in a long-standing lawsuit, new plaintiffs (or members of the plaintiff class) can argue that the remedies fail to vindicate their rights. The loss of educational opportunities, even temporarily, can be irremediable, as many educational experts can attest. Joining ongoing lawsuits provides the additional benefit of helping alter the terms of the debate. If the effort to intervene or break the class in an existing lawsuit is unsuccessful, the advocates subsequently can file a new independent lawsuit.

The advocates also will have to determine whether to proceed with a class action or to proceed on behalf of a group of individual plaintiffs, which may depend on applicable state rules. Class actions have bigger impact: By definition, every member of the class (which can number in the tens of thousands) will be entitled to relief. However, certifying the class presents an additional legal battle, and indeed the sheer numbers in a class action may scare a judge. If school districts are named as defendants, it may be necessary to find class representatives in each district. The advocates will have to perform a careful cost–benefit analysis to decide whether to proceed with a class action or an action on behalf of a group of individuals.

Either way, the lawsuit typically will proceed on behalf of named parents suing on their own behalf and on behalf of their children. It is important to choose dedicated parents who have a compelling story as the lead plaintiffs or class representatives.

Choosing defendants also presents a difficult decision. Failing schools may be scattered across the state. Sweeping all failing schools within the lawsuit makes for a high-impact case and a statewide story. But it can also make for a cumbersome lawsuit. If the lawsuit encompasses multiple districts, the individual districts may be necessary parties, which will result in lots of lawyers on the other side. If the state is primarily responsible for education and its funding, however, it may be

possible to sue the state alone, even if the plaintiffs reside in multiple districts. Alternatively, the advocates may target a small number of especially troubled schools or school districts. Courts may be more willing to grant extraordinary relief if the scope is relatively narrow and confined to school districts that are universally acknowledged to be egregious.

The desired remedy is a pro rata share of the student's education funding to use at a private (or public) school of the family's choice. To the extent that state funds alone are sufficient to cover tuition, that may make it easier to sue the state without the necessity of including school districts. It is important to emphasize such a remedy is not a judicially created voucher program; rather, it is a damages remedy directed at victims of a constitutionally deficient educational system, just as in the IDEA context. Trial experts can also show that choice ensures accountability in the public schools.

Other remedies may be possible depending on local circumstances. In the ongoing New Jersey case, *Crawford v. Davy*, the plaintiffs are seeking both a private school choice remedy and an injunction against residence-based school assignments where they operate to consign children to failing schools. Advocates may also wish to consider alternate remedies, such as lifting caps on charter schools, especially where private school choice may be problematic.

If the advocates are free to choose the venue in which to file, they should do so with an eye toward judges with courage and integrity. Choice-remedy lawsuits should be filed in state courts; federal lawsuits are all but certain to be dismissed unless NCLB is strengthened by adding a private right of action.

Creating the legal team is a crucial decision. Public interest law firms may be particularly adept at prosecuting choice-remedy litigation. Large mainstream law firms can bring useful clout and resources—but they can be expensive. Some may be willing to litigate such cases on a pro bono or discounted basis. Law professors may be willing litigators as well. It is desirable to have a diverse legal team, whose members bring varied experience, backgrounds, political affiliations, and connections. The lead attorney should have sufficient time, expertise, and commitment. Prominent lawyers and law professors can provide credibility by signing on of-counsel.

For most lawyers, the learning curve will be steep. It may be useful to include one or more lawyers who have experience with choice-remedy cases as consultants to the local legal team. Their expertise can help bring the local team up to speed and provide economies of scale in working through the logistical and legal issues and drafting the complaint.

The lawsuit should be coordinated by a well-established nonprofit organization, which can collect tax-deductible contributions for the lawsuit. The organization (and its partners) can take responsibility for community organizing, plaintiff recruitment, data collection and production, media, and political action. The overall team should span the divides of party affiliation, ethnicity, and wealth.



The courtroom efforts should proceed in concert with an aggressive campaign in the court of public opinion and (where feasible) legislative activities. The lawsuit should be announced with a major news conference, and rallies should accompany major court events.

In *Crawford v. Davy*, which I consider a model for choice-remedy litigation, the coordinating roles are provided by Excellent Education for Everyone, the Latino Leadership Alliance, and the Black Ministers Alliance. The lawsuit is extremely well crafted and skillfully guided by two local attorneys, Julio Gomez and Patricia Bombelyn, who in turn are aided by a team of legal advisors. The lawsuit has been covered favorably and extensively throughout the state, fueling legislative efforts to create school choice programs.

The impact of choice-remedy lawsuits can be magnified to the extent that lawsuits in multiple states can be coordinated. Filing one lawsuit is a statewide or regional story; filing two or more is a national story.

Litigation can be a lengthy and grueling process. Investors and participants must gear for a multiyear battle and probable setbacks. But lawsuits can provide a wonderful galvanizing opportunity, especially in states where legislative prospects are dim. Litigation is *action*, which too often is difficult to sustain in states with powerful opposition to school choice. Properly framed, choice-remedy lawsuits can inform and mold public support for school choice while providing an opportunity for tangible progress through judicial or legislative action.

The factors discussed here are likely to arise in all choice-remedy litigation, but there is no magic formula for success. Local circumstances will define the realm of the possible and inform strategy in specific cases. Successive teams of creative lawyers surely will learn from the experiences of their predecessors and improve upon the product. Eventually, with commitment and ingenuity, we will score a litigation breakthrough that will pave the way for additional victories. But along the way, with every step, we will attract to our cause new allies among people of good faith who come through our efforts to recognize the urgency of the problem and the necessity of systemic remedies. In that way, litigation that surely will be difficult to win will nonetheless prove impossible to lose.

## REFERENCES

- Bolick, C. (2003). *Voucher wars: Waging the legal battle over school choice*. Washington, DC: Cato Institute.
- Bolick, C. (2007). *The first line of defense: A blueprint for state constitutional litigation to expand freedom*. Phoenix, AZ: Goldwater Institute.
- Florence County Sch. Dist. No. Four v. Carter by and through Carter*, 510 U.S. 7 (1993).
- Hanushek, E. A. (2006). *Courting failure: How school finance lawsuits exploit judges' good intentions and harm our children*. Stanford, CA: Education Next Books.

- Heise, M. (1995). State constitutions, school finance, and the "third wave": From equity to adequacy. 68 *Temple Law Review* 1151.
- Jenkins v. Leininger*, 659 N.E.2d 1366 (Ill. App. 1995).
- Robinson v. Cahill*, 287 A.2d 187 (N.J. 1973).
- San Antonio Independent School District v. Rodriguez*, 411 U.S. 1 (1973).
- Serrano v. Priest*, 487 P.2d 1241 (Cal. 1971)
- Zelman v. Simmons-Harris* 536 U.S. 639 (2002).

# Hope for Children Trapped in Failing Schools: The Promise of *Crawford v. Davy*

Julio C. Gomez, Esq.  
*Gomez L.L.C., Westfield, New Jersey*

On October 4, 2007, a trial level court in New Jersey dismissed *Crawford v. Davy*, a class action lawsuit filed on behalf of 60,000 schoolchildren throughout the state seeking the court's authority to leave schools that fail to educate their students. By filing suit, plaintiff schoolchildren had hoped to be transferred to an alternative successful public or private school utilizing their pro rata share of state and local school funds to subsidize the transfer. Now, the dismissal of *Crawford* consigns these children to poor inadequate neighborhood schools indefinitely. If the dismissal of *Crawford v. Davy* is not reversed on appeal, it will not only extinguish the hope of plaintiff schoolchildren to receive an equal and adequate educational opportunity, but could threaten the right of a thorough and efficient education guaranteed by the State Constitution and reverse gains achieved over the past 40 years in New Jersey's education jurisprudence. This article places *Crawford* in the context of the state's enduring legal struggle to equalize educational opportunities and discusses its claims and purposes in relation to that history. The article then addresses the significance of the *Crawford* dismissal on the state's legal precedents, especially rulings in the on-going *Abbott v. Burke* equity funding litigation. Finally, the article concludes with a prediction of the impact that *Crawford's* dismissal may pose for the larger equity/adequacy litigation movement playing out across the country. For the moment, the hope of 60,000 plaintiff schoolchildren is diminished. Only time and New Jersey's appellate courts will dictate whether their hope for an equal and adequate education shall survive.

On October 4, 2007, the Chancery Division of the Superior Court of the State of New Jersey dismissed *Crawford v. Davy*, a class action lawsuit filed on behalf of approximately 60,000 schoolchildren throughout the state who seek to leave

---

Correspondence should be sent to Julio C. Gomez of Gomez L.L.C. is a civil rights attorney practicing in Westfield, New Jersey. He is co-counsel for the plaintiffs in *Crawford v. Davy* with Patricia Bombelyn of Perez & Bombelyn P.C., in New Brunswick, New Jersey. E-mail: julio-gomez@comcast.net

schools that fail to educate the majority of their students.<sup>1</sup> Primarily, plaintiffs allege that the schools they are assigned to attend do not provide an equal or adequate educational opportunity and therefore violate their right, under the State Constitution, to receive a thorough and efficient education.

By filing suit, plaintiff schoolchildren hoped to be transferred to alternative successful public (or private) schools utilizing their pro rata share of public school funds to subsidize the transfer. Now, the dismissal of *Crawford* consigns these children to inadequate neighborhood schools indefinitely. If the dismissal of *Crawford* is affirmed on appeal, it will not only extinguish the hope of these children to receive a proper education but could potentially threaten the right to a thorough and efficient education under the State Constitution and begin to reverse gains achieved over the past 34 years in New Jersey's education jurisprudence.

This article discusses the case of *Crawford v. Davy*, the grounds for its dismissal, and the consequences thereof. It begins with a discussion of the state fundamental right to a thorough and efficient education and state litigation to equalize educational opportunities for all children. By placing *Crawford* in that context the legal claims and purposes of the lawsuit are explained. An analysis of the Chancery Court's reasoning and the purported grounds for dismissing *Crawford* follows. The significance of the dismissal on the state right to education, and principally the ongoing *Abbott v. Burke* litigation, is also analyzed. Finally, the article concludes with a prediction about the effect the dismissal of *Crawford* could pose for the larger equity and adequacy litigation explosion blanketing the rest of the country. For the moment, the hope of 60,000 schoolchildren in New Jersey is diminished. Only time and the state's appellate courts will dictate whether their hope for an equal and adequate education can survive.

### CRAWFORD IN CONTEXT: THEMES, CLAIMS, AND PURPOSES

New Jersey has demonstrated a long deep commitment to public education. As early as 1817, the State Legislature enacted a school fund "as a first step toward establishing a state system of public common schools."<sup>2</sup> In 1844 the State Constitution made that fund permanent.<sup>3</sup> Later, in 1875 the State Constitution was amended again to include the well-known language of the Thorough and Efficient

---

<sup>1</sup> See generally *Crawford v. Davy*, Docket No. C-137-06, slip op. and order (N.J. Super. Ct. Ch. Div. Oct. 4, 2007).

<sup>2</sup> Paul L. Tractenberg, *The Evolution and Implementation of Education Rights under the New Jersey Constitution of 1947*, 29 RUTGERS L.J. 827, 832 fn. 17 (1998). However appropriations to support public schools were not authorized until 1829. See *Robinson v. Cahill*, 62 N.J. 473, 506 (1973) (citing I. Myers, *The Story of New Jersey* (1945), pp. 447-450).

<sup>3</sup> N.J. Const. of 1844, Art. IV, §7, ¶6.



(T&E) clause, which required the legislature to “provide for the maintenance and support of a *thorough and efficient* [italics added] system of free public schools” in the State of New Jersey.<sup>4</sup> The T&E clause reappeared in the 1947 (and now current) version of the State Constitution,<sup>5</sup> and has served as the basis for a state fundamental right to education fueling long and tortuous litigation.<sup>6</sup>

In 1895, when the New Jersey Supreme Court interpreted the T&E clause for the first time, it declared that the purpose of the clause “was to impose on the legislature a duty of providing for a thorough and efficient system of free schools, capable of affording *every child* such instruction as is necessary to fit it *for the ordinary duties of citizenship*.”<sup>7</sup> Thus, from its inception, public education under New Jersey law was associated with preparing every person in the state for civic participation, and the court invoked themes of *equality* and *quality* to describe the constitutional mandate. When the State Supreme Court issued its landmark education rulings in the *Robinson v. Cahill* line of cases, the court affirmed and expanded this early reading of the T&E clause reinvoking the same themes:

---

<sup>4</sup>N.J. Const. of 1844, Art. IV, §7, ¶6 (amended 1875). Similar education clauses appear in numerous state constitutions. For example, the “thorough and efficient” language also appears in the constitutions of Maryland, Minnesota, New Jersey, Ohio, Pennsylvania, and West Virginia. See Martin R. West and Paul E. Peterson, *The Adequacy Lawsuit: A Critical Appraisal*, p. 7.

<sup>5</sup>N.J. Const., Art. VIII, §4, ¶1.

<sup>6</sup>The T&E clause has been invoked to uphold a statute permitting free transportation of children to remote public and private schools, see *West Morris Regional Bd. of Educ. v. Sills*, 58 N.J. 464 (1971); to authorize sending students across district boundaries and merging school districts to avoid racial imbalance or segregated schools, see *Jenkins v. Morris Township School District*, 58 N.J. 483 (1971); to direct an increase in a particular school district’s annual school budget to achieve an adequate education, see *Elizabeth Board of Education v. Elizabeth City Council*, 55 N.J. 501 (1970); and to authorize a local board of education to unilaterally alter a collective bargaining agreement to achieve racial diversity among school administrators in response to race riots in the city of Newark. See *Porcelli v. Titus*, 108 N.J. Super 301 (App. Div. 1969). Most notably (and more recently) the clause was invoked to declare an individual fundamental right to an education of a certain quality and to order billions of dollars in increased appropriations for public schools as well as specific school-based policy reforms. See generally *Robinson v. Cahill*, 62 N.J. 473 (1973) and *Abbott v. Burke*, 119 N.J. 287 (1990). It is no surprise therefore that the New Jersey Supreme Court has explicitly acknowledged that “the education of a child has always been of supreme importance and an ideal which has long been required in our State.” *State v. Vaughn*, 44 N.J. 142, 145 (1965).

<sup>7</sup>*Landis v. Ashworth*, 57 N.J.L. 509, 512 (1895) (involving challenge to school tax levied on a local school district). The New Jersey Supreme Court cited the T&E clause in two earlier decisions, *Pierce v. Union District School Trustees*, 46 N.J.L. 76 (1884) (ordering public school to admit Black children under school law entitling all children between the ages of 5 and 18 to free public school), and *Kimball v. Hendee*, 57 N.J.L. 307 (1894) (affirming the status of a de facto board of education, composed of persons actually elected as school trustees at a school meeting, despite action of the county superintendent, in appointing other trustees, upon the supposition that the election was illegally conducted). But *Landis* is the first instance in which the Supreme Court gave meaning to the T&E clause.

We do not doubt that an *equal* educational opportunity for children was precisely in mind. The [constitutional] mandate . . . can have no other import . . . the Constitution's guarantee must be understood to embrace that educational opportunity which is needed in the contemporary setting *to equip a child for his role as a citizen and as a competitor in the labor market* [italics added].<sup>8</sup>

Thus, the concept of an adequate education under the state constitutional mandate involved an equal opportunity to receive not only sufficient preparation for civic duties but also personal vocational development that met the exigencies of modern times. When the State Supreme Court issued another series of landmark rulings, this time in the *Abbott v. Burke* line of cases, this concept of the state right to education was affirmed and described thusly: "At its core, a constitutionally adequate education has been defined as an education that will prepare public school children for a meaningful role in society, one that will enable them to compete effectively in the economy and to contribute and to participate as citizens and members of their communities."<sup>9</sup> The State Supreme Court further added:

The constitutional guarantee of a thorough and efficient education attaches to every school district, and indeed, to every individual school in the State. Of course, the right to a thorough and efficient education does not ensure that every student will succeed. It must, however, ensure that every child in New Jersey has the opportunity to achieve.<sup>10</sup>

In sum, the state right to an education in New Jersey has from its inception required not only an *equal* educational opportunity but an education of a certain kind or *quality*. Even though the State could delegate the task of delivering an education to children, it could not dispense with the duty to achieve the constitutional mandate of educational equity and adequacy. Despite such powerful interpretations of the T&E clause, however, the State Supreme Court has never ordered an immediate remedy to correct a deprivation of a child's right to a thorough and efficient education. Consider *Robinson* and *Abbott*, the state's most prominent education cases.

*Robinson v. Cahill* was an action brought in the early 1970s by residents, taxpayers, and various municipal officials challenging the constitutionality of New Jersey's system of financing public schools. At the time, New Jersey's method of financing public schools relied heavily on local taxation for the bulk of a school district's funding (67%) even though it was clear that certain municipalities did not have sufficient taxable real property to raise enough funds to meet the

<sup>8</sup>*Robinson v. Cahill*, 62 N.J. 473, 513 and 515 (1973) (*Robinson I*).

<sup>9</sup>*Abbott v. Burke*, 149 N.J. 145, 166 (1997) (*Abbott IV*) (citing *Abbott v. Burke*, 100 N.J. 269, 280-81 (1985) (*Abbott I*) and *Robinson v. Cahill*, 62 N.J. 473, 515 (1973)).

<sup>10</sup>*Id.* at 198.

educational needs of their students.<sup>11</sup> Plaintiffs claimed a denial of the right to a thorough and efficient education where local taxation could not achieve levels of school funding equal to other districts. Reasoning that the T&E clause required “equality of educational opportunity,” the Court in *Robinson I*, determined that the financing scheme at the time was unconstitutional and “not demonstrably designed to guarantee that local effort plus the State aid will yield to all the pupils in the State that level of educational opportunity which the 1975 amendment [the T&E clause] mandates.” But, in fashioning an appropriate remedy, the *Robinson* Court fell short. The New Jersey Supreme Court did not issue any immediate remedy for the State’s confirmed failure to comply with the constitutional obligation to provide a thorough and efficient education. Rather, the court heard further argument with respect to appropriate remedies including whether the judiciary could redirect appropriations of the legislature.<sup>12</sup> Following those arguments, the *Robinson II* Court resolved to give the State Legislature a year and half (until December 31, 1974) to adopt revised legislation and specifically withheld any ruling on the consequence of the legislature’s failure to do so.<sup>13</sup> When the legislature failed to heed that deadline, the *Robinson III* Court determined that it would be inequitable to order remedies for the 1975–1976 school year<sup>14</sup> and scheduled still more briefing and oral argument on the scope of its remedial authority and proposed remedies.<sup>15</sup> Four months later, in *Robinson IV*, the Court finally declared that “the right of children to a thorough and efficient education is a fundamental right guaranteed by the [State] Constitution,” and proceeded to order a *redistribution* of approximately \$300,000,000 in state aid funds as a provisional remedy.<sup>16</sup> Thereafter, the state legislature finally enacted a revised funding scheme (the Public School Education Act of 1975) to address the deprivation identified by the Court, and the Court vacated its prior remedy orders. In early 1976, in *Robinson V*, the Court held the new act to be facially constitutional and brought the *Robinson* litigation to a close.<sup>17</sup>

*Robinson* can be described as the first chapter in New Jersey’s saga to equalize and enhance educational opportunities in public schools throughout the state. The *Robinson* line of cases established an individual fundamental right to a thorough and efficient education and the court’s authority to judge violations of that right and to issue remedies to compel enforcement but only after substantial (if not excruciating) deference to the other branches of government to act. Thus, *Robinson*

---

<sup>11</sup>*Robinson v. Cahill*, 62 N.J. 473, 519 (1973) (*Robinson I*).

<sup>12</sup>*Id.* at 520–521. Evidently the Court in *Robinson I* was sensitive to issues of justiciability and separation of powers that judicial review of legislative funding schemes necessarily implicated.

<sup>13</sup>*Robinson v. Cahill*, 63 N.J. 196, 198 (1973) (*Robinson II*).

<sup>14</sup>*Robinson v. Cahill*, 67 N.J. 35, 36–37 (1975) (*Robinson III*).

<sup>15</sup>*Id.* at 37–38.

<sup>16</sup>*Robinson v. Cahill*, 69 N.J. 133, 147 (1975) (*Robinson IV*).

<sup>17</sup>*Id.* at 467.

fell short of establishing any immediate remedy for the violation of a child's right to a thorough and efficient education—a right deemed fundamental. Whether the Public School Education Act would be applied evenhandedly to ensure equal and adequate educational opportunity would have to wait, and thus the stage was set for the second chapter in New Jersey's equity litigation saga, *Abbott v. Burke*.<sup>18</sup>

The case resulting in the landmark *Abbott v. Burke* rulings was initially filed on February 5, 1981.<sup>19</sup> *Abbott* was an action brought by students seeking to declare provisions of the state's school funding statute at that time (the Public School Education Act) unconstitutional on the grounds that it violated the T&E clause. In *Abbott*, the Court declared the act unconstitutional as applied to poorer urban school districts and ordered an amended funding scheme to ensure parity of educational funding between property-rich and property-poor school districts (equity funding) as well as supplemental funding to meet the "special educational needs" of students in property-poor districts (adequacy funding).<sup>20</sup> But like *Robinson*, the *Abbott* Court also deferred to the State Legislature to devise specific amendments to the school funding statute.<sup>21</sup>

Four years later the revised school funding statute (the Quality Education Act) challenged in *Abbott III* was declared unconstitutional because it did not ensure parity between the rich and poor districts and its supplemental funding provisions were not based on any informed study of student needs and real costs.<sup>22</sup> But again, no further remedy was ordered. Three years after that, in *Abbott IV*, another revised school funding statute (the Comprehensive Educational Improvement and Financing Act) was declared unconstitutional because it failed to guarantee sufficient funds for students in poor districts to achieve state standards and supplemental funding was unsupported by any study.<sup>23</sup> This time the Court finally entered a specific order requiring parity funding and a study to determine the amount of supplemental funding that was required and previously ordered among other remedies.<sup>24</sup> A year later, in *Abbott V*, the Court finally approved supplemental funding for a series of programs to meet the "special educational needs" of students in poor districts including whole-school reform, expanded kindergarten and prekindergarten, summer school, school-based health and social services, and other programs.<sup>25</sup>

<sup>18</sup>*Abbott v. Burke*, 100 N.J. 269 (1985).

<sup>19</sup>See *Abbott v. Burke*, 477 A.2d 1278, 1979 (NJ App. Div. 1984) and its progeny.

<sup>20</sup>*Abbott v. Burke*, 119 N.J. 287, 385 (1990) (*Abbott II*).

<sup>21</sup>*Id.* at 388.

<sup>22</sup>*Abbott v. Burke*, 136 N.J. 444, 446–47 (1994) (*Abbott III*).

<sup>23</sup>*Abbott v. Burke*, 149 N.J. 145 (1997) (*Abbott IV*).

<sup>24</sup>*Id.* at 224–26.

<sup>25</sup>*Abbott v. Burke*, 153 N.J. 480, 493 (1998).



Since then the State Supreme Court has rendered several more *Abbott* decisions dealing with a range of issues including but not limited to teacher certification,<sup>26</sup> school building remediation and construction,<sup>27</sup> preschool curriculum and enrollment,<sup>28</sup> and school improvement programs.<sup>29</sup> Further appeals seeking the court's review of the state's implementation of prior *Abbott* Court orders, and the efficacy of those efforts, continues to this day. But consider this: when the lawsuit was filed, young Raymond Arthur Abbott, the lead student-plaintiff in the case, was only 12 years old.<sup>30</sup> Like most complex litigation, *Abbott v. Burke* suffered from time-consuming setbacks and was plagued by appeals; it took 9 years for the New Jersey Supreme Court to review the case on the merits for the first time in *Abbott II* and issue the first decision in favor of the plaintiffs' claims. By that time Raymond Arthur Abbott was 21 years old and had dropped out of high school.<sup>31</sup> Moreover, "despite more than \$3 billion in additional funds" as a result of the *Abbott* decisions, there has been no improvement across the [school] districts that received such funding increases and student achievement in New Jersey's lowest income school districts remains "persistently far worse than that in other school districts in the state."<sup>32</sup> According to Peter Denton, founder and chairman of Excellent Education for Everyone, the most prominent education rights organization in the state, "over the several decades in which New Jersey has tripled spending on its low-income urban schools, their performance has steadily declined, as measured by college attendance rates, standardized test scores, K-12 attendance rates, and high school graduation rates."<sup>33</sup>

Regardless of the extraordinary holdings of the New Jersey Supreme Court in *Robinson* and *Abbott*, the greatest failure of those decisions appears to be the absence of an immediate and effective remedy that directly benefited children, not educational institutions and bureaucrats with increased funds and programs. The effects of increased funding and whole school reforms take years to implement, and if any progress is realized, it occurs long after hundreds of children are sacrificed to trial and error, red tape and incompetence. Today, New Jersey spends more than any other state on K-12 education.<sup>34</sup> Evidently, a new approach was needed to correct deprivations of a child's fundamental right to receive a thorough

---

<sup>26</sup>*Abbott v. Burke*, 163 N.J. 95 (2000); *Abbott v. Burke*, 180 N.J. 444 (2004); *Abbott v. Burke*, 181 N.J. 311 (2004).

<sup>27</sup>*Abbott v. Burke*, 164 N.J. 84 (2000).

<sup>28</sup>*Abbott v. Burke*, 170 N.J. 537 (2002).

<sup>29</sup>*Abbott v. Burke*, 177 N.J. 578 (2003).

<sup>30</sup>Jonathan Kozol, *Savage Inequalities: Children in America's School*, p. 172.

<sup>31</sup>*Id.*

<sup>32</sup>COURTING FAILURE, Eric A. Hanushek, ed., Williamson M. Evers and Paul Clopton, *High-Spending, Low-Performing School Districts*, pp. 133-34.

<sup>33</sup>*Id.*

<sup>34</sup>*Id.* In fact, New Jersey "has been the top spender nearly every year since 1990." *Id.*

and efficient education. Toward the end of its opinion in *Abbott II*, the State Supreme Court made one noteworthy (and somewhat clairvoyant) observation: “*If the children of poorer districts went to school today in richer ones, educationally they would be a lot better off* [italics added].”<sup>35</sup> Thus, the *Abbott* Court ignored the most obvious remedy to the problem it encountered of unequal and inadequate educational opportunity: the immediate transfer of plaintiff schoolchildren from failing schools to good schools.<sup>36</sup> Therein lay the seeds of *Crawford v. Davy*.

*Crawford v. Davy* purports to be the third chapter in New Jersey’s ongoing saga to enforce the state constitutional mandate of an equal and adequate, thorough and efficient, education. *Crawford* was filed precisely to secure once and for all an immediate and meaningful remedy for children who are trapped in schools that fail to educate and do not live up to the standard of thorough and efficient. Plaintiff schoolchildren in *Crawford* do not seek increased funding for their schools, school-based reforms, or supplemental programs. They simply seek the right to leave their assigned school and to attend an alternative school that does not fail the majority of its children, regardless of whether the alternative school is public or private. In New Jersey, as in many other states, children are required to attend their neighborhood school regardless of whether that school complies with state law, has demonstrated an ability to educate its students, or is physically falling apart. The 96 schools listed in the complaint in *Crawford* appear therein because the majority of students in those schools have not been taught the skills and knowledge necessary to pass the state’s basic proficiency examinations. It is the central theme of *Crawford* that no child should be required to attend a school year after year with an ongoing track record of failure. Thus, *Crawford* seeks to correct the shortcomings of the *Robinson* and *Abbott* lines of cases by first and foremost establishing an immediate remedy for the violation of an individual child’s right to a thorough and efficient education that benefits a child directly and by establishing an overwhelming incentive for failing schools (and their districts) to improve outcomes with the threatened loss of their monopoly—the exclusive privilege to educate the children in their neighborhood.

Plaintiff schoolchildren in *Crawford* asserted three distinct legal claims: (a) denial of the right to a thorough and efficient education, (b) denial of the right to equal protection (under both the State and Federal Constitutions), and (c) and a

---

<sup>35</sup>*Abbott v. Burke*, 119 N.J. 287, 394 (1990) (*Abbott II*).

<sup>36</sup>It is not surprising that there is no mention of vouchers or school transfers in the *Abbott* litigation, either in the proceedings at the State Supreme Court or below at the administrative level. Although the plaintiffs were schoolchildren, the type of remedies sought in *Abbott* (primarily equalized funding) inured first and foremost to the benefit of educational bureaucracies that were not providing a thorough and efficient education in the first place—the school districts. A voucher or school transfer remedy would have benefited the economic interests of those institutions far less.

violation of the New Jersey Civil Rights Act.<sup>37</sup> Defendants named in the lawsuit are state officials such as the State Commissioner of Education and the State Board of Education, as well as the 25 local Boards of Education responsible for almost 100 failing schools identified in the complaint.

With respect to the first legal claim—denial of the right to a thorough and efficient education—Plaintiffs allege that they are not provided with the “skills and knowledge they need to pass” the state’s standardized assessment tests; they therefore receive unequal and inadequate “educational opportunities.” The tests that plaintiff schoolchildren do not receive the skills and knowledge to pass are the very tests the state designed to measure attainment of the state’s education standards, denominated Core Curriculum Content Standards (CCCS), which were adopted 10 years ago by the State Department of Education to define the substantive meaning of a thorough and efficient education in New Jersey. Plaintiffs further allege that defendant school boards are “charged with conducting and supervising” their schools “in accordance with constitutional, statutory and regulatory mandates” for public education. Plaintiffs allege the existence of a legal framework of school regulations, state laws, and constitutional mandates that all school boards and state officials are required to follow, including but not limited to, aligning curriculum with CCCS, providing appropriate instruction to underperforming students, and implementing school-level improvement plans. Plaintiffs further allege that defendant state officials must “supervise,” “support,” “review,” “control,” and “enforce” this entire scheme, which they themselves helped to create. Plaintiffs further allege that the law requires *all* students to demonstrate the “knowledge and skills” of CCCS.<sup>38</sup> Because the defendants fail to comply with these requirements, Plaintiffs claim a deprivation of a thorough and efficient education results.

With respect to Plaintiffs’ second legal claim in *Crawford*—the denial of equal protection—Plaintiffs allege that they are similarly situated to other schoolchildren in the state because the State Constitution entitles every school-aged child to a thorough and efficient education and an equal educational opportunity.<sup>39</sup> Plaintiffs

---

<sup>37</sup>See generally *Crawford v. Davy*, Docket No. C-137-06, first amended complaint (N.J. Super. Ct. Ch. Div. Jan. 12, 2007). Technically the complaint in *Crawford* asserts four counts or legal causes of action because the denial of equal protection is alleged separately under the 14th Amendment and the New Jersey State Constitution. *Id.*

<sup>38</sup>Plaintiffs’ Complaint does not encompass every school where not all students demonstrate CCCS; rather Plaintiffs’ Complaint embraces the 96 worst performing schools in New Jersey where the failure to demonstrate proficiency is the norm for the majority, as opposed to minority, of the students.

<sup>39</sup>*Crawford v. Davy*, Docket No. C-137-06, first amended complaint ¶¶61-64, 152 (N.J. Super. Ct. Ch. Div. Jan. 12, 2007); see also *Robinson v. Cahill*, 69 N.J. 133, 147 (1975) (holding that “the right of children to a thorough and efficient system of education is fundamental . . .”); *Abbott v. Burke*, 119 N.J. 287, 296 (1990) (holding children are “constitutionally entitled” to an “equal educational opportunity”).

further allege that defendants treat them differently from other schoolchildren in the state by consigning them year after year to inadequate or failing schools that do not impart the required skills and knowledge that constitute a thorough and efficient education. The defendants consign the plaintiffs to such schools by enforcing district boundaries and residence-based school assignments. District boundaries and residence-based school assignments classify plaintiff schoolchildren on the basis of residence, thereby consigning them to failing schools that deprive scores of children of an equal and adequate education. Plaintiffs further allege that district boundaries and residence-based school assignments do not serve any appropriate governmental objective when they operate to deny plaintiff schoolchildren the education right fundamentally guaranteed by the State Constitution. As result, Plaintiffs allege Defendants treat them unequally by denying to them the same educational opportunities that are afforded to other students in schools that meet the constitutional mandate of thorough and efficient. These actions allegedly constitute a violation of equal protection.

Plaintiffs' third legal claim—violation of the New Jersey Civil Rights Act—simply incorporates each of the first two claims by reference because the act creates a separate statutory cause of action for violations of civil rights.<sup>40</sup> Under the plain meaning of the act deprivations of “equal protection rights” under the Federal Constitution and deprivations of “any substantive rights” under the State Constitution are actionable.<sup>41</sup> As previously stated, plaintiffs asserted a cause of action for denial of equal protection under the federal and state constitutions and a denial of the right to a thorough and efficient education under the State Constitution. Any violation of those civil rights would constitute a violation of New Jersey's Civil Rights Act.

Considering the historical context in which *Crawford* was filed and the small number of potential beneficiaries, the lawsuit is modest by comparison to *Robinson* and *Abbott*. Unlike *Robinson* and *Abbott*, the *Crawford* case was filed after the definition and educational standards for a thorough and efficient education became well defined by statute and regulation, after performance measures for schools had already been developed, and after detailed reports evaluating school district performance were already being released to the public on an annual basis. As a result of such data, only 96 of more than 3,000 schools (in 25 of about 600 school districts) are subject of the suit (roughly 4% of the total student population in New Jersey). In addition, *Crawford* does not seek (or require) any substantial increases in state spending, as Plaintiffs would have public school funds currently expended on their education (their pro-rata share) to be used to fund their transfer to an adequate school. Plaintiffs have also proposed a staggered remedy beginning with

---

<sup>40</sup> See N.J.S.A. 10:6-2(c). The New Jersey Civil Rights Act also provides for reasonable attorney's fees and costs of suit. N.J.S.A. 10:6-2(f).

<sup>41</sup> N.J.S.A. 10:6-2(c).



public school transfers first, followed by private school transfers when capacity in public schools is exceeded, and an out-of-district transfer only when a transfer to a school within district is not possible.

Most notably, *Crawford* was intended to enforce a void that existed at the time the *Robinson* and *Abbott* line of cases were decided: the absence of a regulatory definition of thorough and efficient. In *Robinson* the Court acknowledged the absence of such a measuring stick. Recognizing that “the State ha[d] never spelled out the content of the educational opportunity the Constitution requires,” the *Robinson* Court was left with no alternative but to define educational opportunity in terms of “dollar input per pupil.”<sup>42</sup> Indeed, because the plaintiffs in *Robinson* were seeking to eliminate funding disparities, the *Robinson* Court “was shown no other viable criterion for measuring compliance with the constitutional mandate.”<sup>43</sup> When the first *Abbott* ruling came down, the Court also recognized the absence of an effective substantive definition of a thorough and efficient education and defaulted on funding disparity as the yardstick for measuring constitutional compliance.<sup>44</sup> Two critical limitations resulted from this default yardstick: first, courts were hesitant to apply equal protection analysis to claims of constitutional deprivation fearful of slippery slope concerns that all governmental services (e.g., utilities, social services, security) would be subject to similar equity claims; second, courts did not have a judicially manageable substantive definition of thorough and efficient from which they could assess the appropriateness of funding remedies, or indeed gauge whether increased funding produced the intended result: improved student achievement. *Crawford* sought to move beyond this paradigm. Relying primarily on state educational standards, *Crawford* abandons the funding yardstick for a substantive standard of measure, the statutory and regulatory definition of “thorough and efficient”—not a court-ordered interpretation of an adequate education. Thus, equity in *Crawford* is measured in terms of equal educational opportunities and access to successful schools, not simply dollars, and adequacy is measured in terms of satisfactory student outcomes, namely, achievement of the State’s specific educational standards. Thus, *Crawford* invokes the same themes of *quality* and *equality* that permeate New Jersey’s education jurisprudence, but in different ways.

Finally, *Crawford* was not filed with the intent to dismantle the remedies ordered in the *Robinson* and *Abbott* line of cases. Plaintiffs in *Crawford* are not seeking to reduce the amount of funds that property poor districts receive in state aid or the variety of programs supported by supplemental funds. Plaintiffs in *Crawford*

---

<sup>42</sup>*Robinson v. Cahill*, 62 N.J. 473, 515-516 (1973)(*Robinson I*).

<sup>43</sup>*Id.*

<sup>44</sup>*Abbott v. Burke*, 119 N.J. 287, 317 (1990) (“there is no standard of breadth of curriculum that must be offered, no standard of other commonly accepted educational criteria . . . and no broad-gauged standard of performance of any district”).

are not oblivious to the advantages that equity funding and adequacy programs may present to many children in poor school districts. However, the plaintiffs in *Crawford* insist that such resources belong to schoolchildren and not to the bureaucracies that serve them. To the extent that those bureaucracies fail to educate schoolchildren they must forfeit resources earmarked for those children; they must also forfeit the privilege of providing a service to those children, and the monopoly granted by the state to educate them. The money used to educate children who are underserved should follow those children to better schools. Thus the remedies established in *Abbott*, and those contemplated by *Crawford*, are intended to co-exist—for example, a property poor district deserves parity funding with property rich school districts so that it has the resources to provide an equally adequate educational opportunity; moreover, property poor districts deserve supplemental funding to overcome the adversities that confront the overwhelming majority of the children that they must serve. But, property poor districts do not deserve immunity from accountability; when their schools fail to educate the majority of their children despite increased resources and funding, those students should not be held captive in defective schools. Those students must be transferred (or permitted to transfer out) to better schools. *Crawford* is intended not only to provide an immediate remedy to a child (not the school district—a distinct difference from *Robinson* and *Abbott*) but also to create a powerful incentive for any school district to operate efficiently and efficaciously: the loss of its consumers. Armed with this understanding of *Crawford*, and its place in New Jersey's education jurisprudence, a review of the decision to dismiss the case is now appropriate.

### GROUNDS AND CONTRADICTIONS FOR DISMISSING *CRAWFORD*

Like the *Abbott* case, *Crawford v. Davy* sustained its first setback 6 months ago, when Judge Neil H. Shuster, J.S.C., issued a blistering decision dismissing the lawsuit in its entirety for a number of reasons, but primarily on the grounds that the plaintiffs' claims and the remedy requested involved nonjusticiable political questions.<sup>45</sup> The Court also dismissed the entire case on the alternative grounds

---

<sup>45</sup>See *Crawford v. Davy*, Docket No. C-137-06, slip op. at pp. 27-42 (N.J. Super. Ct. Ch. Div. Oct. 4, 2007). It should not go unnoticed that the Court did not rule against the plaintiffs on the issue of standing. *Id.* at pp. 13-22. The plaintiffs in *Crawford* named 25 local school boards as codefendants with state officials. However, the 15 named representative plaintiffs in *Crawford* attended only 9 of the 25 school districts operated by the defendant school boards. The school boards that operated school districts that none of the 15 named representative plaintiffs attended, therefore, argued for dismissal on the basis that the lawsuit could not proceed as to them without a representative plaintiff from their school district. In essence, such boards had no dealings with the named representative plaintiffs in the complaint and therefore those plaintiffs had no legal "standing" to assert claims against them. Plaintiffs

that the plaintiffs had not pled a cause of action with respect to the denial of the right to a thorough and efficient education and the denial of equal protection.<sup>46</sup> In addition, the Court held that even if the claims were justiciable, and properly pled, that the plaintiffs were required to exhaust their administrative remedies before seeking relief in a court of law.<sup>47</sup> Finally, the Court dismissed the lawsuit as to the 25 local school board defendants on the grounds that such defendants lacked legal authority to provide plaintiff schoolchildren with any of the remedies they sought.<sup>48</sup>

The issue of political question nonjusticiability “is primarily a function of [the] separation of powers” doctrine.<sup>49</sup> Our system of government is based on the principle that the powers of government shall be divided among three distinct branches (the legislative, executive, and judicial) and that no one branch shall exercise the powers properly belonging to either of the others. To decide whether a matter is justiciable, a court must determine not only whether it is authorized to review the matter but also whether judicially identifiable and judicially manageable standards exist to render a decision. The court dismissing the lawsuit concluded that *Crawford* was not justiciable because it raised matters that were committed to another branch of government, the State Legislature, and that there were no judicially discoverable or judicially manageable standards for rendering a decision. According to the court, there were no standards to determine if a violation of plaintiffs’ rights had occurred and no standards to issue a remedy in the form of a school transfer or voucher. The court reached this conclusion by reasoning that the issues in *Crawford* were textually committed to the legislature under the T&E clause; that the legislature’s role in education in New Jersey is fundamental and primary; and that “in the absence of constitutional or statutory standards, it is not the function of [the] Court to substitute its judgment for that of the Legislature with respect to the rules it has adopted or the procedures followed in giving effect to the constitutionally-declared scheme.”<sup>50</sup> According to the court:

Plaintiffs seek to have the Court devise and adopt a standard for determining when the fundamental right to a “thorough and efficient” education is in fact being deprived, rather than have the Court follow an already existing framework for determining this

---

countered by arguing that the 15 named representative plaintiffs could represent children in those other school districts because the claims and issues would be similar (if not identical) and, furthermore, that all school boards are legally related or “juridically linked” to the state defendants as agents who carry out a uniform policy that is depriving plaintiff schoolchildren of their civil rights. The Court agreed and held that the 15 named representative plaintiffs had legal standing to sue 25 local school boards. *Id.* at 22.

<sup>46</sup>*Id.* at 42–48.

<sup>47</sup>*Id.* at 48–50.

<sup>48</sup>*Id.* at 22–27.

<sup>49</sup>See *Baker v. Carr*, 369 U.S. 186 (1962).

<sup>50</sup>*Crawford v. Davy*, Docket No. C-137-06, slip op. at p. 30 (N.J. Super. Ct. Ch. Div. Oct. 4, 2007).

issue. Such a decision is clearly non-justiciable. Moreover, Plaintiffs seek to have the Court order that consecutive years of failing Assessment scores constitutes “failing” to provide a “thorough and efficient education” . . . [T]he Court finds it lacks the ability to “judicially determine” that consecutive years of failing Assessment scores, alone, constitutes a “breach” of the “duty” to provide a “thorough and efficient education,” and that it lacks the authority to “judicially mold” a remedy to “protect” that duty. Essentially, there is a lack of judicially discoverable and manageable standards for determining the “breach.” Moreover, determination of such issues are “impossibl[e]. . . without an initial policy determination of a kind clearly for nonjudicial discretion; or the impossibility of a court’s undertaking independent resolution without expressing lack of respect due coordinate branches of government.”

The court’s reasoning hereabove is flawed for a number of reasons. First, *Crawford* does not require any review of the action of the State Legislature at all. No statute or legislative action is challenged by the plaintiffs in *Crawford*—a situation quite unlike the *Robinson* and *Abbott* lines of cases where plaintiffs challenged the legislature’s manner of funding public school education and the State Supreme Court reviewed that scheme, held it unconstitutional, and established parameters for the legislature to follow in re-devising it.<sup>51</sup> Even if *Crawford* required a court to review legislative action, the State Supreme Court in *Robinson* specifically considered whether such involvement by a court would violate the separation of powers doctrine and whether such claims would raise a nonjusticiable political question.<sup>52</sup> Rejecting that argument the State Supreme Court reasoned as follows:

The people in 1875 ordained the Legislature to be their agent to effectuate an educational system but did not intend to tolerate an unconstitutional vacuum should the Legislature default in seeing to their specification that the system be thorough and efficient. We have adjudicated such a default. Under emerging modern concepts as to judicial responsibility to enforce constitutional right there has been no paucity of examples of affirmative judicial action towards such ends.<sup>53</sup>

The State Supreme Court further reasoned,

The argument is recast in terms of the doctrine of separation of powers, purportedly precluding judicial direction for expenditure of State moneys, that being exclusively for the other Branches. . . . The interest here at stake transcends that of an ordinary individual claimant against the State. It is that of all the school children of the

<sup>51</sup>See e.g. *Robinson v. Cahill*, 69 N.J. 133, 150 (1975) (*Robinson IV*) (ordering a redistribution of \$300,000,000 in school funding appropriated by the state legislature); *Abott v. Burke*, 149 N.J. 145, 198 n. 35 and 223 (1997) (ordering the state legislature to increase funding for 28 school districts by upwards of \$248,000,000).

<sup>52</sup>*Robinson v. Cahill*, 69 N.J. 133, 151-155 (1975) (*Robinson IV*).

<sup>53</sup>*Id.* at 152.



State, guaranteed by the constitutional voice of the sovereign people: equality of educational opportunity. This Court, as the designated last-resort guarantor of the Constitution's command, possesses and must use power equal to its responsibility. Sometimes, unavoidably incident thereto and in response to a constitutional mandate, the Court must act, even in a sense seem to encroach, in areas otherwise reserved to other Branches of government.<sup>54</sup>

As a result, the State Supreme Court has already determined that the kind of inquiry required by plaintiffs' claims in *Crawford* is in fact justiciable. But, even though the plaintiffs argued this precedent to the Court, no consideration of it appears anywhere in the Court's opinion dismissing *Crawford*. The Court simply makes no attempt to reconcile its decision with this precedent from *Robinson*.

Second, no fair reading of the complaint in *Crawford* supports the view that plaintiffs "seek to have the Court devise and adopt a standard for determining when the fundamental right to a thorough and efficient education is in fact being deprived," or that the Court not "follow an already existing framework," as the dismissal opinion states. On the contrary, as discussed above, *Crawford* is entirely based on an existing statutory and regulatory framework that requires the state (and local boards) to identify educational standards that define the meaning of a "thorough and efficient" education, to administer tests that measure student achievement of those educational standards, to set uniform proficiency benchmarks demonstrating adequate progress towards achieving those standards, to report the results of those tests, to review the performance of schools and school districts using the percentage of students performing proficiently as a measure, and to provide appropriate instruction to improve the skills and knowledge for students performing below established levels of student proficiency. The existing framework commands in no uncertain terms that "[a]ll students shall be expected to demonstrate the knowledge and skills of the CCCS as measured by the Statewide assessment [test] system."<sup>55</sup> The framework also provides that a school district "may be certified" if it achieves the state's proficiency benchmarks on these examinations as "providing a thorough and efficient system of education."<sup>56</sup> Clearly, demonstrating proficiency on the state's assessment tests is a *prerequisite* to a thorough and efficient education. In view of this explicit statutory and regulatory framework, which plaintiffs embraced fully and cited extensively in the complaint and in their briefs, it is quite stunning that the Court takes the position that plaintiffs "seek to have the Court devise and adopt [its own] standard" and "not follow an already existing framework." The point of *Crawford* is that a legal framework *already* exists but is not being followed, either by state officials or local school

---

<sup>54</sup>*Id.* at 154.

<sup>55</sup>N.J.A.C. 6A:8-4.3(d).

<sup>56</sup>See N.J.A.C. 6A:8-4.4(c)(1) and N.J.S.A. 18A:7A-14.

boards. The Court dismissing *Crawford* inexplicably appears to deny the existence of that framework.

Moreover, the existing framework would give the Court test and year specific standards to judicially determine each of the issues it claims to lack standards for. With respect to determining that consecutive years of failing test scores *alone* constitutes a “breach” of the “duty” to provide a “thorough and efficient education,” the framework specifically provides judicially discoverable and manageable standards in the form of proficiency percentage benchmarks (which are actually more rigorous than the ones employed by plaintiffs).<sup>57</sup> A school district must meet those benchmarks to be certified by the state as providing a thorough and efficient education.<sup>58</sup> Similarly, the Court could employ those very same benchmarks and, based on the evidence presented at trial, determine whether they are being met or not by the defendants. Relying on the regulatory requirement that a school district may be certified as providing a thorough and efficient education once those benchmarks are met, the Court could easily rule that schools not meeting those benchmarks are not providing a thorough and efficient education. The Court need not develop or adopt any standard of its own to make these judicial determinations. The Court might even be able to entertain defenses from the defendants to excuse their failure to meet the required benchmarks, if the framework provides for such defenses. But again, the matter would be justiciable nonetheless.

With respect to the court’s claimed lack of authority to “judicially mold” a remedy to “protect” a breach of the duty to provide a thorough and efficient education, the Court need only follow the specific example of the State Supreme

---

<sup>57</sup>See generally N.J.A.C. 6A:8-4.4. In *Crawford* Plaintiffs employ an average uniform standard to evaluate school performance. Plaintiffs allege that any school that achieves proficiency of only 49% or less on both the Language Arts and Mathematics assessments fails to provide a thorough and efficient education and any school that achieves proficiency of merely 24% or less on either the Language Arts or Mathematics assessment fails to provide a thorough and efficient education. New Jersey’s regulatory standards are more grade and year specific and more rigorous than the standard employed by plaintiffs to plead their case. For example, among fourth graders, schools and school districts were required to achieve 68% language arts proficiency and 53% mathematics proficiency in 2003–2004, 75% language arts proficiency and 62% mathematics proficiency from 2004 to 2007, and 82% language arts proficiency and 73% mathematics proficiency by the current academic year, 2007–2008. N.J.A.C. 6A:8-4.4(a)(1)(i). Among eighth graders, schools and school districts were required to achieve 58% language arts proficiency and 39% mathematics proficiency in 2003–2004, 66% language arts proficiency and 49% mathematics proficiency from 2004 to 2007, and 76% language arts proficiency and 62% mathematics proficiency by the current academic year, 2007–2008. N.J.A.C. 6A:8-4.4(a)(2)(i). Similarly, among high school students, schools and school districts were required to achieve 73% language arts proficiency and 55% mathematics proficiency in 2003–2004, 79% language arts proficiency and 64% mathematics proficiency from 2004 to 2007, and 85% language arts and 74% mathematics proficiency by the current academic year. N.J.A.C. 6A:8-4.4(a)(3)(i). None of the schools identified in *Crawford* fully comply with these proficiency percentage benchmarks; rather they perform abysmally below these regulatory standards.

<sup>58</sup>See N.J.A.C. 6A:8-4.4(c)(1) and N.J.S.A. 18A:7A-14.

Court in the *Abbott* line of cases. For instance, in *Abbott II* and *Abbott III*, the State Supreme Court declared that “supplemental” funding was necessary to address the social and economic disadvantages of children in poor urban districts and ordered it for that purpose. The State having failed to act appropriately on supplemental funding by the time of *Abbott IV*, the Supreme Court ordered as follows:

The determination of appropriate remedial relief in the critical area of the special needs of at-risk children and the programs necessary to meet those needs is both fact-sensitive and complex; it is a problem squarely within the special expertise of educators. A court alone cannot, and should not, assume the responsibility for independently making the critical educational findings and determinations that will be the basis for such relief. *We can, however, provide necessary procedures and identify the parties who best may devise the educational, programmatic, and fiscal measures to be incorporated in such remedial relief.* Accordingly, we remand the matter to the Superior Court to implement that aspect of the Court’s remedial order.

The Superior Court, consistent with this opinion, shall direct the Commissioner to initiate a study and to prepare a report with specific findings and recommendations covering the special needs that must be addressed to assure a thorough and efficient education to the students in the SNDs [special needs districts]. That report shall identify the additional needs of those students, specify the programs required to address those needs, determine the costs associated with each of the required programs, and set forth the Commissioner’s plan for implementation of the needed programs. In addition, the Superior Court shall direct the Commissioner to consider the educational capital and facility needs of the SNDs and to determine what actions must be initiated and undertaken by the State to identify and meet those needs.

The parties shall be given the opportunity to participate in the proceedings conducted by the Commissioner and to respond to and file exceptions to the Commissioner’s report prior to its submission to the Superior Court.

*The Superior Court may, in addition, conduct hearings with the participation of the Commissioner and all parties. The Superior Court may appoint, with the approval of this Court, a Special Master to assist the court in all proceedings and in reaching its determinations and rendering its decision.* The Superior Court, based on its review of the Commissioner’s report, any additional evidence, and any findings and determinations of the Special Master, shall render a decision with its findings, conclusions, and recommendations covering the special programs that should be implemented in the special needs districts and the costs of their implementation. That decision will be made available to all parties, and shall be reviewed by this Court. [italics added]<sup>59</sup>

The preceding Order of the State Supreme Court in *Abbott IV* succinctly describes the power that a court in New Jersey has to formulate the remedy plaintiffs are seeking in *Crawford*. Upon ordering that a school transfer is necessary to

---

<sup>59</sup> *Abbott v. Burke*, 149 N.J. 145, 199-201 (1997)(*Abbott IV*).

correct deprivations of the state constitutional right to an education, a court could order a study to develop a plan to effectuate school transfers. In its subsequent decision, *Abbott V*, the State Supreme Court further commanded as follows: "We, therefore, direct the Commissioner to promulgate regulations and guidelines that will codify the education reforms incorporated in the Court's remedial measures."<sup>60</sup>

Likewise, in *Crawford*, a court could order the Commissioner to promulgate regulations and guidelines to affect transfers from failing schools. Considering what the State Supreme Court actually held and ordered over the past 16 years in *Abbott*, the court dismissing *Crawford* clearly has legal authority to judicially mold and order an appropriate remedy by identifying the cause of the constitutional deprivation and requiring the Commissioner to study and develop a specific remedy that corrects it. In their briefs (and at oral argument), Plaintiffs specifically proposed such a procedure.<sup>61</sup> But the court dismissing *Crawford* renounced its authority to mold a remedy in this fashion. By doing so, it deviated from the *Abbott* precedent and arguably committed reversible legal error. In sum, by concluding that *Crawford* presents non-justiciable political questions, the court dismissing *Crawford* ignores the statutory and regulatory legal framework that provides judicially identifiable and manageable standards to rule in the case, and ignores or misreads well-established precedents that authorize it to mold a judicial remedy.<sup>62</sup>

The court's grounds for dismissing *Crawford* for plaintiffs' failure to plead violations of the rights to a thorough and efficient education and equal protection are equally flawed and subject to reversal. Regarding plaintiffs' failure to plead a violation of the right to a thorough and efficient education, the court appears to

---

<sup>60</sup>*Abbott v. Burke*, 153 N.J. 480, 526 (1998)(*Abbott V*).

<sup>61</sup>*Crawford v. Davy*, Docket No. C-137-06, Plaintiffs' omnibus memorandum of law in opposition to all dispositive motions filed by Defendants at p. 47 (N.J. Super. Ct. Ch. Div. January 31, 2007) (This Court could always issue a declaratory judgment regarding the constitutional violations alleged by the plaintiffs' complaint; restrain the enforcement of district boundaries only when *applied* to consign plaintiff schoolchildren to failing schools; and then submit the "voucher" remedy to the Commissioner for specific study and recommendation just as our State Supreme Court did [with supplemental aid programs] in *Abbott IV*).

<sup>62</sup>It should not go unnoticed that the court's opinion dismissing *Crawford* fails to discuss (and arguably) consider the holdings of certain landmark U.S. Supreme Court precedents cited by the plaintiffs on the issues of justiciability and a court's broad remedial authority to correct violations of civil rights, namely, *Baker v. Carr*, 369 U.S. 186 (1962) (striking down state's General Assembly apportionment statute because it diluted or debased the black vote in violation of equal protection), *Gomillion v. Lightfoot*, 364 U.S. 339 (1960) (applying the 15th Amendment to strike down a redrafting of municipal boundaries that affected a discriminatory impairment of voting rights despite "sweeping commitment" to state legislatures of the power to draw and redraw such boundaries), and *Swann v. Charlotte-Mecklenburg Bd. of Educ.*, 402 U.S. 1, 15 (1971) (where school board failed in its duty to devise a desegregation plan, court had authority to appoint its own expert and impose a plan).



apply the same reasoning it did on the issue of nonjusticiability. Because the court denies any existing framework that requires schools to achieve a certain level of proficiency on the state's assessment tests as a prerequisite for demonstrating a thorough and efficient education, then presumably a deprivation of that right cannot be plead on that basis. Regarding plaintiffs' failure to plead a violation of equal protection (whether under the State or Federal Constitution), the court adopts the unsupportable position that plaintiffs are required to allege (and prove) that school district boundaries were intended by the legislature to deprive some children of an equal and adequate education when they were enacted. The former may constitute reversible legal error for the same reasons stated earlier: a statutory and regulatory framework *already* exists that requires a certain level of proficiency to demonstrate compliance with the constitutional mandate of thorough and efficient. The latter may constitute reversible legal error because there is no legal requirement to allege (much less prove) intent to discriminate when a facially neutral law is applied to deprive a fundamental right to any group of persons. Again, turning to the *Robinson* and *Abbott* line of cases, the State Supreme Court dealt with facially neutral funding schemes. But in *Robinson*, the Court never required plaintiffs to prove that the legislature intended to deprive an equal educational opportunity to students in school districts with low real property rates. In *Abbott*, the Court also never required plaintiffs to prove that the legislature intended to deprive a thorough and efficient education to students in poor urban school districts. Similarly, in *Crawford*, the Court dismissing the suit should not have required the plaintiffs to plead that the legislature intended municipal school boundaries to deprive plaintiffs of an equal and adequate education.

The court's grounds for dismissal of the local boards may also constitute legal error. The dismissal of all local school boards from *Crawford* is premised on the theory that such boards "only have authority to act within the statutory framework within which they were created."<sup>63</sup> Because the state does not give local school boards "authority to simply ignore district boundaries or compulsory attendance laws," then such boards "cannot unilaterally provide the relief sought by Plaintiffs,"<sup>64</sup> no relief may be obtained from them, and they are not proper defendants in the case. However, the court's reasoning completely disregards that plaintiffs have requested relief in the form of a transfer to any successful public school. To the extent that a successful public school exists within the same school district in which plaintiffs reside, then defendant school boards would not have to cross district boundaries in order to effectuate a remedy. The court's dismissal of the local boards is evidently premised upon the ill-conceived notion that plaintiffs are only seeking interschool district transfers in reassigning a child to an adequate school.

---

<sup>63</sup>*Crawford v. Davy*, Docket No. C-137-06, slip op. at 23 (N.J. Super. Ct. Ch. Div. Oct. 4, 2007).

<sup>64</sup>*Id.* at 25.

Finally, the dismissal of *Crawford* on the alternative grounds of Plaintiffs' failure to exhaust administrative remedies may also be reversible, even though it may present a closer legal question. The doctrine of exhaustion of administrative remedies requires a prospective plaintiff to seek relief in all administrative for a prior to filing suit in court. Judicial efficiency is its primary purpose. In New Jersey, the doctrine may require the plaintiffs in *Crawford* to file an administrative complaint with the State Department of Education and participate in hearings before an administrative law judge whose decision would be reviewed by the Commissioner of Education (a defendant in *Crawford*) and later by the State Board of Education (yet another defendant in the suit). On this issue, the court dismissing *Crawford* holds that "no court should decide constitutional issues in a vacuum, in the absence of a well-developed record isolating the essential factual question at their basis and including findings of fact." As a result, the court concludes that the plaintiffs' claims are "premature in the absence of well-developed factual record before the "special expertise" of Defendants' agency heads, and if appropriate, the Legislature."<sup>65</sup> Plaintiffs in *Crawford* therefore must exhaust the administrative remedy of prosecuting an administrative complaint at the Department of Education before coming to the court for redress. In reaching its conclusion, however, the court fails to consider (and therefore distinguish) any of New Jersey's exceptions to the exhaustion requirement that are cited at length by the plaintiffs (e.g., when only questions of law are at issue; when administrative remedies are futile; when irreparable harm would result; when the jurisdiction of the agency is doubtful; or when an overriding public interest calls for a prompt judicial resolution, exhaustion of administrative remedies is *not* required). More important, the court fails to consider plaintiffs' argument that the ultimate decision makers in any administrative proceeding (the Commissioner and State Board of Education) have already prejudged the case based on the arguments they presented in their briefs to dismiss the case. Therefore, requiring plaintiffs to seek administrative relief essentially from the defendants themselves is a *fait accompli* and utterly futile. Plaintiffs in *Crawford* are not oblivious to the fact that the State Supreme Court ordered the plaintiffs in *Abbott* to exhaust administrative remedies before pursuing their claims in court.<sup>66</sup> However, that experience bears out plaintiffs' futility argument:

In the *Abbott* line of cases, where plaintiffs were ordered to exhaust their so-called administrative "remedy," the Administrative Law Judge declined to rule on remedies, the Commissioner of Education declined to accept the recommendations made by the Administrative Law Judge (including findings of fact *and* conclusions of law),

<sup>65</sup>*Id.* at 48 and 50.

<sup>66</sup>*See Abbott v. Burke*, 100 N.J. 269 (1985).

and the State Board of Education “adopted the Commissioner’s decision in almost all respects.” *Abbott v. Burke*, 119 N.J. 287, 297-300 (1990). The plaintiffs in *Abbott* lost 4 years to that process.<sup>67</sup>

As plaintiffs argued in their brief, “what justice can plaintiff schoolchildren expect to receive if they are ordered by th[e] Court to plead their case directly to the State Defendants and to ask them to declare themselves in violation of the Constitution?”<sup>68</sup> Ultimately, the administrative process can produce no record that cannot also be produced in the first instance in a court of law, giving the state defendants every opportunity to present their expertise, and giving all parties the additional procedural safeguards afforded by the Rules of Evidence and the Rules of Civil Procedure, not to mention a speedier resolution of matters concerning a substantial public concern—the proper education of 60,000 schoolchildren.

In sum, the Court’s decision to dismiss *Crawford* on all of the aforementioned bases is replete with unsupportable conclusions that are contrary to established precedent. Apart from failing to address in its opinion a multitude of the plaintiffs’ arguments and citations to controlling precedent, the Court also misapplies the cases it does cite. Ironically, by failing to acknowledge the existing legal framework, the Court actually engages in an analysis that is based less on judicial standards and more on judicial prejudices. Consider the following remarkable commentary in the Court’s opinion: “The Court questions how it could adequately safeguard what Plaintiffs would suggest are “successful” schools from becoming “failing” schools, if the Court were to permit a mass exodus of approximately 60,000 schoolchildren.”<sup>69</sup>

Without hearing any evidence at all this Court has apparently decided that allowing the plaintiff schoolchildren in *Crawford* to transfer to better schools would only cause the schools receiving them to fail. Evidently, the Court has predetermined that the plaintiff schoolchildren should be blamed for the failure of their schools, not the teachers or administrators who are responsible for, and paid to, educate those children.

The dismissal of *Crawford* should be reversed on appeal simply because it does not follow well-established legal precedent. If it is not reversed, however, the opinion dismissing *Crawford* may pose grave consequences for the individual right to a thorough and efficient education in the state, the ongoing *Abbott* litigation, and perhaps the larger equity/adequacy legal movement throughout the country.

---

<sup>67</sup>*Crawford v. Davy*, Docket No. C-137-06, plaintiffs’ omnibus surreply to all dispositive motions filed by defendants at p. 27 (N.J. Super. Ct. Ch. Div. April 6, 2007).

<sup>68</sup>*Id.*

<sup>69</sup>*Crawford v. Davy*, Docket No. C-137-06, slip op. at p. 38 (N.J. Super. Ct. Ch. Div. Oct. 4, 2007).

CONSEQUENCES OF *CRAWFORD*'S DISMISSAL

The consequences of *Crawford*'s dismissal could conceivably diminish the right to a thorough and efficient education under the New Jersey Constitution; it could also reverse gains perceived in the *Robinson* and *Abbott* lines of cases, as well as direct the course of the ongoing equity/adequacy litigation in a manner ill-suited to securing enhanced educational opportunities for children across the country.

With respect to the right to a thorough and efficient education as guaranteed by the New Jersey Constitution, the court's holding clearly weakens it. First, the holding essentially renounces the existing statutory and regulatory framework that substantively defines that right. Since the *Landis* case in 1895, the State Supreme Court has struggled to define the meaning of a thorough and efficient education. In both *Robinson* and *Abbott* the Court explicitly acknowledged the failure of the legislative and executive branches to define that right. But the state's adoption of substantive educational standards (in the form of CCCS) was held to be "a major step to spell out and explain the meaning of a constitutional education."<sup>70</sup> Not surprisingly the State Legislature enacted statutes requiring continuous review and revision of the state's educational standards and the Department of Education promulgated regulations requiring ongoing review and readoption every 5 years.<sup>71</sup> Moreover, the standards themselves explicitly state that they "define what all students should know and be able to do by the end of their public school education."<sup>72</sup> It could be argued that the current framework is the product of several contributing forces, such as the State's deeply rooted commitment to education; the legislative and executive response to the judiciary's findings and orders in *Robinson* and *Abbott*; not to mention the educational standards movement, the requirements of the federal No Child Left Behind Act, and the influence of the teacher's union (the New Jersey Education Association).

For these reasons, the Court's decision to refuse to acknowledge that framework as a basis for asserting a violation of the right to a thorough and efficient education is particularly egregious. The current framework did not develop overnight and certainly involved extensive consideration and negotiation over many years by many stakeholders and each of the three branches of government. It is not worthy of the short shrift given to it in the opinion dismissing *Crawford*. Moreover, it is not clear from the Court's opinion whether the Court simply failed to grasp the significance of the existing framework in relation to the issues in the case or

<sup>70</sup>*Abbott v. Burke*, 149 N.J. 145, 167-68 (1997) (*Abbott IV*).

<sup>71</sup>N.J.A.C. 6A:8-2.1. The review process requires "advisory panels of public school educators, higher education representatives, business representatives, and other citizens" to recommend revised education standards, preapproval publication of any proposed standards and public hearings before final approval by the State Board of Education.

<sup>72</sup>N.J.A.C. 6A:8-1.1(a).



whether the Court regards the existing framework as only a set of aspirations or goals that cannot be legally enforced. The Court also fails to demonstrate how a deprivation of the right to a thorough and efficient education should be pled, regardless of the existing regulatory framework. One is left to wonder whether a deprivation of the right can be pled at all.

If *Crawford's* dismissal is not reversed, the opinion could be cited in the future for the proposition that the right to a thorough and efficient education can only be understood in terms of funding and that no substantive definition of the right is controlling. The implication would be that no claim of deprivation could be based on substantive inequality regarding deficient teaching methodologies, disparate teacher qualifications, unaligned curricula, or student underperformance. The opinion could be cited more broadly, however, for the proposition that any alleged deprivation of the right to a thorough and efficient education is non-justiciable by a court of law because the legislature has primary (and arguably exclusive) responsibility under the State Constitution to maintain and enforce a thorough and efficient system of public schools. It could be cited to argue that there are simply no judicially discoverable or manageable standards for determining a violation of the state right to an education or formulating an appropriate remedy. Such an application of the opinion dismissing *Crawford* would eviscerate the constitutional mandate. Of what benefit to anyone is a judicially unenforceable fundamental right? Chief Justice John Marshall aptly stated in the landmark decision of *Marbury v. Madison*, "it is a general and undisputable rule, that where there is a legal right, there is also a legal remedy, by suit or action at law, whenever that right is invaded."<sup>73</sup> Not so, according to *Crawford's* dismissal.

Remarkably, the *Abbott* litigation has not concluded and the implications that *Crawford's* dismissal could pose for *Abbott* are considerable. Most recently, in *Abbott*, plaintiffs have sought further review of the state's implementation of various orders issued in prior *Abbott* decisions.<sup>74</sup> But the State Supreme Court has deemed those claims premature and ordered plaintiffs to recast their arguments in light of the budget for fiscal year 2008. The dismissal of *Crawford* could serve as a basis for the State Supreme Court to retreat from the perpetual oversight of *Abbott* remedies. Worse, the dismissal opinion could be used to encourage the Court to revisit the separation of powers doctrine and the conflict posed by judicial review of State compliance with the constitutional mandate, urging reversal of the court's rulings in *Robinson* and *Abbott*. As indicated earlier, the dismissal of *Crawford* on grounds of nonjusticiability is irreconcilable with the *Robinson* and *Abbott* lines of cases. If the dismissal of *Crawford* stands, then *Robinson* and *Abbott* should be reversed.

<sup>73</sup>*Marbury v. Madison*, 5 U.S. (1 Cranch) 137, 163 (1803).

<sup>74</sup>*Abbott v. Burke*, 2007 WL 1518909, slip op. (N.J. May 24, 2007).

Finally, the dismissal of *Crawford* may impact the larger equity/adequacy legal movement. At last count, more than 125 cases have been filed throughout the country challenging the constitutionality of public school funding schemes. Previously, these plaintiffs sought increased funding for disadvantaged school districts (in *equity* suits akin to the *Robinson* line of cases); currently, the trend has moved toward overall increases in school funding to achieve a certain level of educational quality (in *adequacy* suits more like *Abbott*).<sup>75</sup> The breadth and scope of this legal movement is staggering:

Beginning in the 1990s, enactment in virtually every state of learning objectives and curriculum standards provided a new reference point for plaintiffs arguing that funding was inadequate overall. By 2006, the constitutionality of funding mechanisms in 39 states had been challenged on adequacy grounds. Indeed, through the first half of 2006, funding mechanisms in only five states – Delaware, Hawaii, Mississippi, Nevada, and Utah – have been spared constitutional challenge.<sup>76</sup>

Scholars posit that “the national push for educational standards and accountability” has fueled such litigation, noting that once “several states moved quickly on their own to establish proficiency standards and regular assessments of the performance of their students,” then “plaintiffs in adequacy cases soon began citing newly collected data on student proficiency, which routinely revealed student performance to be lagging well below state targets.”<sup>77</sup> In that respect, *Crawford* does not differ. New Jersey’s educational standards and state assessments play *the* central role in Plaintiff’s legal theory in *Crawford*. The dismissal of *Crawford*, however, suggests that state educational standards are not judicially enforceable, however, even if they were passed specifically to define a constitutional right or obligation of the State. Like *Crawford*, any adequacy suit that uses state standards to define an adequate education would fail. But the import of the *Crawford* dismissal is much graver than a failed funding suit. The *Crawford* dismissal stands for the proposition that achievement or failure to meet state educational standards can only be gauged by school boards and administrators who are less likely to police themselves and hold each other accountable. The dismissal also compels the view that educational standards themselves may not be legal requirements but rather aspirational goals, unenforceable in any court of law. Either way, accountability in public education is reduced and parents will encounter greater difficulty ensuring that educational delivery systems actually meet the needs of their children.

---

<sup>75</sup>James W. Guthrie and Matthew G. Springer, *Courtroom Alchemy, Adequacy Advocates Turn Guesstimates Into Gold*, EDUCATION NEXT (Winter 2007), p. 21.

<sup>76</sup>*Id.*

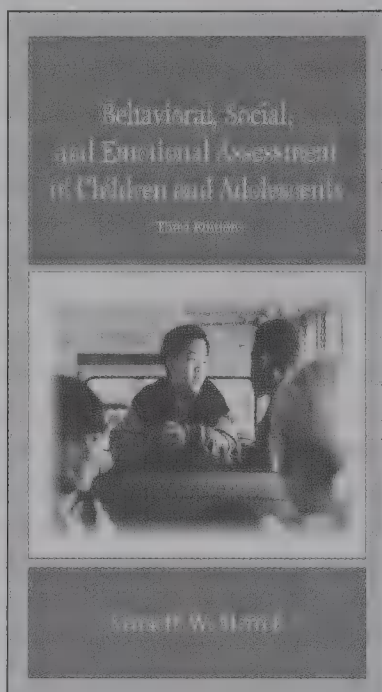
<sup>77</sup>Martin R. West and Paul E. Peterson, *The Adequacy Lawsuit: A Critical Appraisal*, p. 8.

## CONCLUSION

*Crawford v. Davy* promises to give children access to better schools when they are assigned to deficient schools that do not educate their students. By ignoring well-established precedent the dismissal of *Crawford v. Davy* weakens a child's right to an equal and adequate education in the State of New Jersey (and possibly elsewhere). The dismissal could serve to deprive litigants of a substantive definition of their right upon which to base their renewed claims of equity and adequacy in education. The dismissal could also persuade courts to abandon their role as "last-resort guarantors" of constitutional rights, depriving plaintiff schoolchildren of a forum in which to seek an immediate and meaningful remedy. The gravest implication of *Crawford's* dismissal is that state educational standards may not be judicially enforceable and that a court does not have authority to oversee compliance with a constitutional mandate to educate. In other words, courts may not consider deprivations of the constitutional right to an education or issue any remedy to correct an alleged deprivation. If the dismissal of *Crawford* is affirmed on appeal, schools that fail to educate their students in New Jersey will likely continue to fail. If the children in such schools are not allowed to leave, then their future participation as citizens will be compromised and their productivity as competitors in the marketplace will be disadvantaged. Without standards, without a court to enforce them, and without an immediate remedy, scores of children will remain trapped in failing schools year after year. *Crawford v. Davy* must be reversed, and hope restored.

# A Fundamental Resource for Your Child & Adolescent Assessment Course!

"A **CLEARLY-WRITTEN, INTERESTING** and **COMPREHENSIVE** guide to the social/emotional assessment of children and adolescents." —*NASP COMMUNIQUE\**



## 3RD EDITION NOW AVAILABLE!

August 2007 • 7 x 10 • 568 pp. • 978-0-8058-5370-4 • \$79.95

In its new 3rd edition, *Behavioral, Social, and Emotional Assessment of Children and Adolescents* continues to be the most comprehensive book in its field as a primary text for the basic assessment course in school psychology curricula. By providing an all-inclusive foundation for conducting clinical assessments of child and adolescent social-emotional behavior in a practical, scientific, and culturally appropriate manner, it is aimed at graduate students, practitioners, and researchers in the fields of school psychology, child clinical psychology, and special education.

### NEW TO THIS EDITION:

- Expanded coverage of ethnic, cultural, and linguistic diversity within socio-emotional assessment.
- New material on conducting assessments within the context of ecological theory and public health models.
- Increased information on assessing children with Asperger's disorder or high-function Autism.
- Discussions of several new assessment tools and systems.
- Recent updates to professional standards and organizational ethics code of APA and NASP.

"This book is...**THE PREMIER VOLUME AVAILABLE** on the behavioral, social, and emotional assessment of school-aged children and adolescents." —*PSYCHOLOGY IN THE SCHOOLS\**

"It presents important information related to the assessment of the behavioral, social, and emotional functioning of children and adolescents in a manner that is understandable to novices yet still valuable for experienced practitioners...**AN EXCELLENT TRAINING RESOURCE** for graduate students and school personnel as well as a valuable reference for practitioners providing services to children and adolescents."

—*JOURNAL OF PSYCHOEDUCATIONAL ASSESSMENT\**

\*Praise for earlier editions

 **Routledge**  
Taylor & Francis Group  
an **informa** business

**ORDERING OPTIONS**  
Mail to: 7625 Empire Drive, Florence, KY 41042  
Call toll free: 1-800-634-7064 • Call international: 859-525-2230  
Fax toll free: 1-800-248-4724 • Fax international: 859-647-5027

 **Lawrence Erlbaum Associates**  
Taylor & Francis Group  
an **informa** business

Please visit [www.routledge.com](http://www.routledge.com) for more information on these and other Education books published by the Taylor & Francis Group.



## INSTRUCTIONS FOR AUTHORS

**Content:** *Peabody Journal of Education (PJE)* publishes quarterly symposia in the broad area of education, including but not limited to topics related to formal institutions serving students in early childhood, preschool, primary, elementary, intermediate, secondary, postsecondary, and tertiary education. The scope of the journal includes special kinds of educational institutions, such as those providing vocational training or the schooling for students with disabilities. PJE also welcomes immediate framework of institutions, and education matters that are important to nations outside the United States. Finally, it includes topics that are linked to the social and organizational context in which formal and informal education take place. The Editor cooperates with groups of scholars to present multifaceted, integrated expositions of important topics. A given issue of PJE may contain contributions from social scientists, historians, philosophers, attorneys, practitioners, and policymakers. Unsolicited proposals for special issues - including designation of participating scholars and an outline of articles - will be accepted for review. Additionally, the Editor cooperates with Editorial Board members to identify potential topics, guest editors, and contributors. PJE has the flexibility to consider publishing monographs or a series focused on particular lines of inquiry. In all cases, the Editor and the Editorial Board will ensure that each issue is carefully reviewed and its articles will comprise a high-quality contribution to understanding and practice.

**Manuscript Preparation:** To submit a prospectus for a special issue, note submission requirements presented here and contact the Coordinating Editor, Michele Thompson, for complete details (address below). Prepare manuscripts according to the *Publication Manual of the American Psychological Association* (5th ed., 2001; available from APA Order Department, P.O. Box 2710, Hyattsville, MD 20874), especially with regard to reference lists and text citations. Follow Guidelines to Reduce Bias in Language (*APA Manual*, pp. 61-87). Manuscripts should include the theme title and rationale for the theme. Using 8.5 x 11-in., nonsmear paper and 1-in. margins, type all components double spaced and in the following order: title page (p. 1), abstract (p. 2), text (including quotations), references, appendices, footnotes, tables, and figure captions. On page 1, type article title, author name(s) and affiliation(s), running head (abbreviated title, no more than 45 characters and spaces), author notes and acknowledgments, and name and address of person to whom requests for reprints should be addressed; on page 2, type an abstract of no more than 150 to 200 words. Indent all paragraphs. Use footnotes sparingly. Attach photocopies of all figures. Number all manuscript pages (including photocopies of figures).

**Cover Letter, Permissions, Credit Lines:** In a cover letter, include the contact author's complete mailing address, e-mail address, and telephone and fax numbers. State that the manuscript includes only original material that has not been published and that is not under review for publication elsewhere. Authors are responsible for all statements made in their work and for obtaining permission to reprint or adapt a copyrighted table or figure or to quote at length from a copyrighted work. Authors should write to original author(s) and original publisher to see if permission is required and to request nonexclusive world rights in all languages to use the material in the current article and in future editions. Include copies of all permissions and credit lines with the manuscript. (See p. 175 of the *APA Manual* for samples of credit lines.)

**Manuscript and Computer Disk Submissions:** Prepare manuscript on a word processor and submit three (3) high-quality printouts and a disk to the Coordinating Editor, Michele Thompson, *Peabody Journal of Education*, Department of Leadership, Policy, and Organizations, Peabody College of Vanderbilt University, Peabody #41, 230 Appleton Place, Nashville, TN 37203-5721. The disk should contain two files - word processor and ASCII versions of the manuscript - and file content must match the printed manuscript exactly, or there will be a delay in publication. Manuscripts and disks are not returned.

**Accepted Manuscripts:** Authors are strongly encouraged to submit manuscripts on disk. The disk should be prepared using MS Word or WordPerfect and should be clearly labeled with the authors' names, file name, and software program. A hardcopy printout that exactly matches the disk must be supplied. Each manuscript must be accompanied by a statement that it has not been published elsewhere and that it has not been submitted simultaneously for publication elsewhere. Authors are responsible for obtaining permission to reproduce copyrighted material from other sources and are required to sign an agreement for the transfer of copyright to the publisher. All accepted manuscripts, artwork, and photographs become the property of the publisher.

All parts of the manuscript should be typewritten, double-spaced, with margins of at least one inch on all sides. Number manuscript pages consecutively throughout the paper. Authors should also supply a shortened version of the title suitable for the running head, not exceeding 50 character spaces. Each article should be summarized in an abstract of not more than 100 words. Avoid abbreviations, diagrams, and reference to the text in the abstract.

**Illustrations:** Illustrations submitted (line drawings, halftones, photos, photomicrographs, etc.) should be clean originals or digital files. Digital files are recommended for highest quality reproduction and should follow these guidelines:

- 300 dpi or higher
- Sized to fit on journal page
- EPS, TIFF, or PSD format only
- Submitted as separate files, not embedded in text files

Color illustrations will be considered for publication; however, the author will be required to bear the full cost involved in their printing and publication. The charge for the first page with color is \$900.00. The next three pages with color are \$450.00 each. A custom quote will be provided for color art totaling more than 4 journal pages. Good-quality color prints should be provided in their final size. The publisher has the right to refuse publication of color prints deemed unacceptable.

**Tables and Figures.** Tables and figures (illustrations) should not be embedded in the text, but should be included as separate sheets or files. A short descriptive title should appear above each table with a clear legend and any footnotes suitably identified below. All units must be included. Figures should be completely labeled, taking into account necessary size reduction. Captions should be typed, double-spaced, on a separate sheet. All original figures should be clearly marked in pencil on the reverse side with the number, author's name, and top edge indicated.

**Proofs and Reprints.** Page proofs are sent to the designated author using Taylor & Francis' Central Article Tracking System (CATS). They must be carefully checked and returned within 48 hours of receipt. Reprints of individual articles are available for order at the time authors review page proofs. A discount on reprints is available to authors who order before print publication.

# Peabody Journal of Education

Volume 83, Number 2, 2008

Introduction to the Special Issue on From Equity to Adequacy to Choice <i>Michael Podgursky</i>	171
A New Defendant at the Table: An Overview of Missouri School Finance and Recent Litigation <i>Michael Podgursky, James Smith, and Matthew G. Springer</i>	174
What Do Cost Functions Tell Us About the Cost of an Adequate Education? <i>Robert Costrell, Eric Hanushek, and Susanna Loeb</i>	198
Can Judges Improve Academic Achievement? <i>Jay P. Greene and Julie R. Trivitt</i>	224
Spending Money When It Is Not Clear What Works <i>Paul T. Hill</i>	238
Next Needed Steps in the Evolution of American Education Finance and Policy: Attenuating a Judicially Imposed Policy Distraction, Activating a Balanced Portfolio of K-12 School Reforms, Advancing Rationality as a Goal in Pursuing Productivity, Advocating Change in a Responsible and Effective Manner <i>James W. Guthrie</i>	259
A Framework for Choice Remedy Litigation <i>Clint Bolick</i>	285
Hope for Children Trapped in Failing Schools: The Promise of <i>Crawford</i> <i>v. Davy</i> <i>Julio C. Gomez, Esq.</i>	297



0161956X(2008)83(2)

ISBN 080589226-5



9 780805 892260

9000



Volume 83 Issue 3 ISSN 0161-956X 2008

# *Peabody Journal of Education*

CURRENT YR/VOL

Marygrove College Library  
8425 West McNichols Road  
Detroit, MI 48221

Published in cooperation with the Peabody College of Vanderbilt University



Routledge  
Taylor & Francis Group



# PEABODY JOURNAL OF EDUCATION

Camilla P. Benbow, Dean, Peabody College  
James W. Guthrie, Editor  
Michele Thompson, Coordinating Editor  
Deborah M. Enright, Associate Editor

## EDITORIAL ADVISORY BOARD

Darlyne Bailey, *University of Minnesota*  
R. Dale Ballou, *Peabody College of Vanderbilt University*  
Sir Clive Booth, *Oxford Brookes University, England*  
Mary M. Brabeck, *New York University*  
Mark Bray, *The University of Hong Kong, PRC*  
Louis A. Castenell, Jr., *The University of Georgia*  
David Y. Chen, *Chongquin Technology and Business University (CTBU)*  
Mike Dishman, *McLocklin, Murphy & Dishman, L.L.P.*  
Lora Cohen-Vogel, *Florida State University*  
Joseph P. Du Cette, *Temple University*  
Victor Firsov, *Russian Academy of Teachers Retraining, Russia*  
Juan Froemel, *LLECE/UNESCO, Chile*  
Louis M. Gomez, *Northwestern University*  
Gerardo M. Gonzalez, *Indiana University*

David W. Gordon, *Elk Grove Unified School District, California*  
Eric A. Hanushek, *Hoover Institution, Stanford University*  
Stephen P. Heyneman, *Peabody College of Vanderbilt University*  
Phillip W. Jones, *The University of Sydney, Australia*  
Kazuo Kuroda, *Waseda University, Japan*  
James H. Lytle, *Trenton Public Schools, New Jersey*  
Michael K. McLendon, *Peabody College of Vanderbilt University*  
Ann L. Marcus, *New York University*  
Bill McKinney, *Region IV Education Service Center, Texas*  
Francois Orivel, *University of Bourgogne, France*  
Penelope L. Peterson, *Northwestern University*  
Michael J. Podgursky, *University of Missouri-Columbia*  
Robert Slavin, *Johns Hopkins University*  
Jandhylala B. G. Tilak, *National Institute of Education Planning and Administration, India*

---

**Abstracted and/or Indexed in:** PsycINFO/Psychological Abstracts; Linguistics and Language Behavior Abstracts; Contents Pages in Education; Sociological Abstracts; Education Index; Education Abstracts; Social Services Abstracts; EBSCOhost Products; Cabell's Directories; Family Index Database.

*Peabody Journal of Education* (ISSN: 0161-956X) is published quarterly in February, May, August, and November for a total of 4 issues per year by Taylor and Francis Group, LLC, 325 Chestnut Street, Suite 800, Philadelphia, PA 19106.

**US Postmaster:** Please send address changes to Peabody Journal of Education, Taylor & Francis Group, LLC, 325 Chestnut Street, Suite 800, Philadelphia, PA 19106.

### Annual Subscription, Volume 83, 2008

Print ISSN - 0161-956X, Online ISSN - 1532-7930

Institutional subscribers: US \$462, UK £277, €370

Personal subscribers: US \$58, UK £35, €46

Institutional and individual subscriptions include access to the online version of the journal. Institutional subscriptions include access for any number of concurrent users across a local area network. Individual subscriptions are single username/password only.

**Production and Advertising Office:** 325 Chestnut Street, Suite 800, Philadelphia, PA 19106. Tel: 215-625-8900, Fax: 215-625-8563.  
Production Editor: Michelle McCarthy

### Subscription offices

**USA/North America:** Taylor and Francis Group, LLC, 325 Chestnut Street, Suite 800, Philadelphia, PA 19106. Tel: 215-625-8900, Fax: 215-625-2940.

**UK/Europe:** Taylor and Francis Customer Services, Sheepen Place, Colchester, Essex CO3 3LP, UK. Tel: +44 (0)-20-7017-5544, Fax: +44 (0)-20-7017-5198.

For a complete guide to Taylor and Francis Group, LLC's journal and book publishing programs, visit our website: [www.taylorandfrancis.com](http://www.taylorandfrancis.com)

**Copyright © 2008 Taylor and Francis Group, LLC.** All rights reserved. No part of this publication may be reproduced, stored, transmitted, or disseminated in any form or by any means without prior written permission from Taylor and Francis Group, LLC. Taylor and Francis Group, LLC grants authorization for individuals to photocopy copyright material for private research use. The publisher assumes no responsibility for any statements of fact or opinion expressed in the published papers. The appearance of advertising in this journal does not constitute an endorsement or approval by the publisher, the editor, or the editorial board of the quality or value of the product advertised or of the claims made for it by its manufacturer.

**Permissions.** For further information, please visit: <http://www.tandf.co.uk/journals/permissions.html>

August 2008



# Introduction to the Special Issue on Scaling Up Teaching and Learning Improvement in Urban Districts: The Promises and Pitfalls of External Assistance Providers

Amanda Datnow  
*University of California, San Diego*

Meredith I. Honig  
*University of Washington*

This special issue of *Peabody Journal of Education* brings together educational scholars across disciplines to examine two significant and related trends in urban school districts: efforts to scale-up high-quality teaching and learning districtwide and the role of external assistance providers in the process. The time is ripe for this issue.

As many have noted, school district leaders across the country are launching prominent initiatives to strengthen teaching and learning. These efforts move beyond pilots that focus on pockets of students and aim to improve teaching and learning for all students often as part of ambitious educational equity and achievement agendas. These developments are fueled by a number of factors. Among them, persistently disappointing results with more limited, programmatic approaches to improvement not centrally focused on teaching and learning have prompted some district leaders to make districtwide teaching and learning improvement the core goal and main focus of their investments. No Child Left Behind and state and local accountability initiatives increase the urgency for districts to increase the scope and depth of their efforts at producing demonstrable gains in academic achievement for all students. Superintendents and other district central

office leaders are coming to understand the importance of their participation in educational improvement strategies not solely when it comes to operations and management but also with regard to learning-focused leadership (e.g., Hightower, Knapp, Marsh, & McLaughlin, 2002; Hubbard, Stein, & Mehan, 2006).

Despite these trends fueling the development of districtwide teaching and learning improvement initiatives, efforts to implement these initiatives have been fraught with challenges, particularly for urban school districts. For example, researchers have documented that implementation of these efforts typically suffers from a lack of clarity among district leaders and staff regarding teaching and learning goals and an absence of enough stable and capable personnel to foster new, promising practices throughout district systems (e.g., Corcoran, Fuhrman, & Belcher, 2001; Hubbard et al., 2006). Some districts struggle to put new teaching and learning initiatives into action in the wake of multiple other sometimes conflicting change initiatives (Hatch, 2001).

Perhaps as a consequence, many urban districts are not going it alone. Some have begun to establish close partnerships with external organizations to assist them in their districtwide teaching and learning improvement efforts. These organizations—such as school reform support organizations, universities, and others—promise to bring a host of new knowledge-based, social, fiscal, and other resources beyond what schools and districts would be able to marshal on their own. When they operate as intermediary organizations they bring these resources to bear not only on schools or district central offices but throughout district systems (Honig, 2004). What do we know about these organizations?

A growing number of studies suggest that, under certain circumstances, these organizations do realize their promise of building district capacity for meeting ambitious teaching and learning goals. For example, a cross-national study of districts in Canada, the United States, and England that have brought about effective change in their schools specifically identified their relationships with external partners as among the conditions that were essential to making such improvements possible (Fullan, Bertani, & Quinn, 2004). Galucci and colleagues have shown how partnerships between the Center for Educational Leadership at the University of Washington and districts in Washington State and nationwide seem associated with changes in teachers' and principals' practice (Gallucci & Boatright, 2007; Gallucci & Swanson, 2008).

However, as Smylie and Corcoran (2006) pointed out in a review of research on these organizations, external assistance providers vary in their ability to help. A host of internal and external circumstances may bolster their work but also lead to their own internal instability and otherwise frustrate the assistance they aim to provide in districts.

Smylie and Corcoran's (2006) conclusions also underscore that these organizations are fundamentally dependent on others for realizing their own goals in ways that also present challenges. We think of their work as a moving mosaic in which

the external providers interact with multiple teachers, administrators, and others in highly complex ways that change over time. Whether the outcomes of external organizations are “successful” depends not just or even mainly on the external organizations themselves but also on those others with whom they engage in their work (Datnow, Hubbard, & Mehan, 2002; Honig, 2004).

Educational scholarship has only just begun to tap these and other complexities related to how external assistance providers operate, the impacts associated with their work, and the conditions that help or hinder them in their efforts, particularly when it comes to districtwide improvements in teaching and learning. Many important questions still remain. Among them, how is the work of external providers organized? How do external providers collaborate with teachers and school and district leaders to improve their capacity to support teaching and learning improvement? What appear to be more and less productive ways of external providers working with educators to produce changes in teaching and learning? How do organizational, political, institutional, and other factors shape the work of external providers and the implementation of new strategies in district central offices, schools, and classrooms?

The articles in this special issue address these questions with theoretically and empirically rich analyses of nationally prominent external assistance providers and their work with urban schools and districts. Each article focuses on elaborating the types of assistance relationships that seem to support teaching and learning initiatives and implications for research and practice in urban settings.

First, the article by Honig and Ikemoto examines how the Institute for Learning (IFL) at the University of Pittsburgh has partnered with multiple urban districts to assist them in developing their capacity for strengthening teaching and learning districtwide. These authors draw on ideas from sociocultural learning theory to identify key features of the IFL-district assistance relationships that seem associated with particular capacity-building outcomes. Overall, they frame the IFL-district relationships as “adaptable assistance relationships.” With this term they underscore how the IFL-district relationships involved (a) IFL staff and district leaders working together to co-construct district capacity-building strategies and (b) the IFL’s efforts to continually revise and refine their work with districts over time as the IFL increased their own knowledge about how to build district capacity and as circumstances in their partner districts evolved and changed.

The article by Coburn, Bae, and Turner also examines how external assistance providers and district leaders co-construct or negotiate their work over time. They use frame analysis and sensemaking theory to examine the work of an external support provider in a midsize urban school district. They explain that such partnerships between districts and external providers hold promise for realizing district goals but are challenging when it comes to defining roles in implementation and managing authority relationships and status differentials. Their findings elaborate that the codesign of efforts to improve teaching and learning is often the result

of complex interactions between members of the external organization, district practitioners, and organizational and political realities. They identify the local and broader policy conditions under which such partnerships are likely to be more and less successful.

The article by Datnow and Park examines how Success for All (SFA) creates and builds knowledge for school improvement. The SFA program is an interesting case to study because of its high-profile success in scaling up to many urban schools and its reputation as a particularly prescriptive comprehensive school reform model. Datnow and Park explain that the theory, strategy, and tools driving the SFA Foundation's approach to school reform seem technically oriented and highly prescribed. However, the deeper process by which the SFAF created knowledge for school improvement in the two featured urban schools was highly dynamic. Their work involved collaboration, negotiation, and conflict along several dimensions—relationships between schools and the SFAF, local and state contexts, and federal educational policy. The interconnections among these dimensions shaped SFA Foundation's strategies toward knowledge development and, in turn, SFA Foundation influenced the educational policy landscape and the school reform process.

The article by Marsh, Hamilton, and Gill examines how Edison Schools, a for-profit educational management organization, combines structured assistance for teaching and learning with accountability mechanisms to achieve school improvement. Drawing on data from a four-year study of Edison Schools, the authors identify the factors that facilitated or inhibited the implementation and outcomes of the work of this external assistance provider. They address issues such as the strength of instructional leadership, the role of union-district relations, and the broader accountability context in shaping implementation results.

Finally, the article by Supovitz takes a step back from the in-depth examinations of the prior pieces and explores the place of external partnerships in districtwide instructional reform efforts. Drawing on data collected in a multiyear study of a large urban school district, Supovitz argues that there are some functions for which districts are essential, others for which they are interchangeable with other types of support organizations, and still other functions for which they are fundamentally inadequate. Findings reveal there are service tasks that are best done by districts in partnership with external support organizations, who have greater resources and expertise to develop the necessarily sophisticated (in design, if not enactment) curricular and training interventions for teachers. However, school support cannot be completely outsourced to intermediary organizations, as districts must assume the essential role of orchestrating the many components of instructional reform that coalesce in schools and classrooms.

Taken together, the articles in this special issue provide theoretically and empirically informed understandings concerning how external support providers work together with educators in schools and districts in the implementation of complex



districtwide teaching and learning improvement initiatives. We hope they help fuel many conversations and investigations into how these partnerships can strengthen the capacity of district systems to realize ambitious teaching and learning improvement goals for all students.

## REFERENCES

- Corcoran, T., Fuhrman, S. H., & Belcher, C.L. (2001). The district role in instructional improvement. *Phi Delta Kappan*, 83(1), 78–84.
- Datnow, A., Hubbard, L., & Mehan, H. (2002). *Extending educational reform: From one school to many*. London: Routledge Falmer.
- Fullan, M., Bertani, A., & Quinn, J. (2004). *Leading in tough times: New lessons for districtwide reform*. Center for Development and Learning. Retrieved July 13, 2007, from [http://www.cdl.org/resource/library/articles/leading\\_in\\_tough\\_times.php?type=subject&id=36](http://www.cdl.org/resource/library/articles/leading_in_tough_times.php?type=subject&id=36)
- Gallucci, C., & Boatright, B. (2007, February). *Gaining traction through professional coaching: A partnership between the center for educational leadership and highline school district* (Interim Research Report 2). Seattle: Center for the Study of Teaching and Policy, University of Washington. Retrieved March 15, 2007, from <http://depts.washington.edu/uwcel/resources/papers.html>
- Gallucci, C., & Swanson, J. (2008, January). *Balancing direction and support—third year scale up of a system-wide instructional improvement reform: The partnership between the Center for Educational Leadership and Norwalk-LaMirada Unified School District* (Interim Research Report 2). Seattle: Center for the Study of Teaching and Policy, University of Washington. Retrieved January 31, 2008, from <http://depts.washington.edu/uwcel/resources/papers.html>
- Hatch, T. (2001). Incoherence in the system: Three perspectives on the implementation of multiple initiatives in one district. *American Journal of Education*, 109, 407–437.
- Hightower, A. M., Knapp, M. S., Marsh, J. A., & McLaughlin, M. W. (Eds.). (2002). *School districts and instructional renewal*. New York: Teachers College Press.
- Honig, M. I. (2004). The new middle management: Intermediary organizations in education policy implementation. *Educational Evaluation and Policy Analysis*, 26, 65–87.
- Hubbard, L., Stein, M. K., & Mehan, H. (2006). *Reform as learning: School reform, organizational culture, and community politics in San Diego*. New York: Routledge.
- Smylie, M., & Corcoran, T. (2006, April). *Nonprofit organizations and the promotion of evidence-based practice*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

# Adaptive Assistance for Learning Improvement Efforts: The Case of the Institute for Learning

Meredith I. Honig  
*University of Washington*

Gina S. Ikemoto  
*RAND Corporation*

Districts across the country face significant demands to strengthen student learning districtwide, and many are turning to intermediary organizations to help them build their capacity for such demanding, large-scale work. However, how these “learning-support intermediary organizations” assist with these capacity-building efforts is little understood. This article reports data from a largely qualitative investigation into how one such intermediary organization, the Institute for Learning (IFL) at the University of Pittsburgh, partnered with multiple urban districts to help build district capacity for districtwide learning improvements. Our conceptual framework draws on sociocultural learning theory to identify key features of the IFL-district assistance relationships that seem associated with these outcomes. We utilized data from interviews, observations, document reviews, and focus groups conducted over a five-year period. Findings elaborate specific features of their assistance relationships—which we call adaptive assistance relationships—such as enabling particular forms of modeling, tools, and opportunities for rich dialogue. We conclude with implications for the research and practice of districtwide learning improvement efforts and the participation of intermediary organizations in the process.

---

Research for this article was funded by the Spencer Foundation and the John D. and Catherine T. MacArthur Foundation as part of the “Meta Study on the Relationship between Research and Practice.” We thank Cynthia Coburn and Mary Kay Stein, Principal Investigators of the Meta-Study, for their feedback on various drafts of this and related articles. Thank you to Lauren Resnick for her careful reading of our work and all the staff of the Institute for Learning who welcomed us in (one of us for many years) to learn with them about their experiences.

Correspondence should be sent to Meredith I. Honig, College of Education, University of Washington, Miller Hall Box 353600, Seattle, WA 98195. E-mail: mihonig@u.washington.edu

Within the last 15 years, school districts have come under increasing pressure to enhance their capacity to play key leadership roles in strengthening student learning districtwide, and many have called on intermediary organizations to help them in the process (Coburn, 2001; Honig, 2004). For example, local organizations such as the Center for Educational Leadership at the University of Washington partner with districts overtime to coach school in strengthening student learning and, in tandem, help district leaders shift their own practice to provide such assistance themselves (Gallucci, Boatright, Lysne, & Swinnerton, 2006). In the 1990s, design teams participating in the New American Schools initiative worked between schools and their central offices to support implementation of whole-school improvement strategies (Berends, Bodilly, & Kirby, 2002; Datnow, Hubbard, & Mehan, 2002). We call these organizations “learning-support intermediaries” because they focus their work specifically on supporting learning improvements and because they occupy a distinct position between central offices and schools where they aim to leverage changes at both levels (Honig, 2004). What more specifically are the activities of learning-support intermediary organizations that seem associated with increased district capacity for learning improvements? Education policy researchers have only begun to elaborate distinct roles for intermediary organizations and their particular contributions to districtwide learning improvement efforts.

We aim to add to this emerging research base by drawing on over more than five years of data on the Institute for Learning (IFL)—which is housed at the University of Pittsburgh—and its learning-support relationships with eight urban districts across the country. The IFL provided a strategic opportunity for this inquiry because it has sustained partnership relationships with multiple districts over almost a decade, specifically around building school- and central-office capacity for strengthening student learning districtwide and because the IFL seems to have contributed to demonstrable improvements in district capacity in a number of respects. The IFL also had access to various fiscal and intellectual resources typically in short supply in external support organizations (Bodilly, Glennan, Kerr, & Galegher, 2004). Accordingly, the IFL promised to demonstrate intermediary-district partnerships functioning at a reasonably high level. Our data come from 264 interviews, more than 232 hr of observations, and focus group and archival data. We used sociocultural learning theory to analyze our data because it promised to help reveal features of assistance relationships associated with deepening practitioners’ engagement in complex work practice.

We found that our conceptual framework captured major dimensions of the IFL-district assistance relationships including certain forms of modeling, tools, and social opportunities. We call these relationships “adaptive assistance relationships” to emphasize their dynamic and locally constructed nature. We draw on the IFL case to suggest future directions for the research and practice of districtwide

learning improvement efforts and the participation of intermediary organizations in the process.

## INTERMEDIARY ORGANIZATIONS AND DISTRICTWIDE LEARNING IMPROVEMENT EFFORTS

Research on districtwide efforts to strengthen student learning is a relatively recent literature that has begun to mushroom within the last 15 years (e.g., Corcoran, Fuhrman, & Belcher, 2001; Elmore & Burney, 1998; Hightower, Knapp, Marsh, & McLaughlin, 2002; Hubbard, Mehan, & Stein, 2006; Murphy & Hallinger, 2001; Snipes, Doolittle, & Herlihy, 2002; Spillane et al., 2002; Spillane & Thompson, 1997). This research seems to converge on a few key findings about these efforts, and we used these findings to frame our research questions. For one, the engagement of central office administrators, school principals, and teachers in student learning improvement is itself a problem of learning—a challenge of supporting professional learning throughout district systems (e.g., D. K. Cohen, 1982; McLaughlin, 2006). For example, from a cognitive learning perspective and across multiple districts, Spillane and colleagues have demonstrated that how school principals and central office administrators make sense of complex reform demands significantly mediates their responses to those demands (Spillane, 2000, 2002; Spillane et al., 2002). Hill, Coburn, and others have elaborated these findings with examinations of teachers and school principals (Coburn, 2001, 2005; Hill, 2001). The literature on teacher professional learning communities also demonstrates that particular forms of collaboration among teachers fosters teacher learning in ways that can bolster teachers' professional practice and student learning (e.g., McLaughlin & Talbert, 2001; Scribner, Cockrell, Cockrell, & Valentine, 1999; Scribner, Hager, & Warne, 2002). Some researchers (e.g., Leithwood & Louis, 1998) have argued that continuous improvement processes that engage teachers and principals, sometimes called "organizational learning," are essential to learning improvements. Examinations of San Diego school's efforts to foster professional learning across all levels of the district reinforce such findings (e.g., Hubbard et al., 2006).

Two, supporting professional learning represents nontraditional work for many district practitioners. Accordingly, even those with strong political will and significant resources to support professional learning struggle with implementing professional learning support systems. For example, Corcoran et al. (2001) chronicled how efforts to strengthen student learning in three districts were curbed by various conditions including uncertainty among central office administrators regarding how they could participate productively in the implementation of such efforts. Hubbard et al.'s (2006) multiyear examination of San Diego's "reform as learning" revealed that teachers, principals, and central office administrators



alike struggled with limited knowledge of and experience with the new work practices that learning improvement initiatives aimed to promote. Some suggest that districtwide improvement efforts fundamentally demand that leaders “manage ambiguity” (Honig, 2001) or “learn to lead what they don’t yet know” (Swinerton, 2006) and that supports for such learning-on-the-job can be few and far between.

Three, some external organizations seem to offer important assistance with these professional learning processes. For example, the Center for Educational Leadership mentioned earlier provides key assistance mainly to school principals and teachers in strengthening their professional practice around reading and literacy. Stein and Brown’s examination of QUASAR also revealed how external organizations bring coaching resources to districts that seem to contribute to demonstrable learning improvements (Stein & Brown, 1996). These and related studies suggest that certain external organizations may help improve district capacity for learning improvements. However, they also show that these organizations are associated with these improvements at only a handful of participating schools absent central-office reform or engagement of professionals throughout district systems in particular work practices (Berends, Chun, Schuyler, Stockly & Briggs, 2002a, 2002b; Bodilly & Berends, 1999; Datnow, Hubbard, & Mehan, 2002; Kirby et al., 2002). These findings suggests that a particular type of external assistance provider may be important to districtwide learning improvements—intermediary organizations or organizations that work between levels of hierarchical district systems (e.g., between classrooms, teachers and principals, principals and central office administrators) to engage practitioners at all levels in deepening their work practice in support of student learning (Honig, 2004). What do learning-support intermediary organizations do when they seem to strengthen district capacity for districtwide learning improvements?

## CONCEPTUAL FRAMEWORK<sup>1</sup>

Building on the emerging literature on reform as learning, we turned to ideas from sociocultural learning theory to ground our inquiry into the participation of intermediary organizations in these processes. This strand of learning theory seemed particularly appropriate to this inquiry because it elaborates what deepening professional work practice entails and how assistance relationships matter to such professional development.

By many accounts, sociocultural learning theory stems from the work of Vygotsky and his students and colleagues such as Leont’ev. These scholars explored how learning unfolds not through an individual’s acquisition of information

---

<sup>1</sup>This conceptual framework is adapted from Honig (in press).

but through an individual's engagement with others and various artifacts or tools in particular social, historical, and cultural contexts (Vygotsky, 1978). Through such engagements, learners socially construct the meaning of particular ideas and in the process develop and also potentially shape the habits of mind of their cultures (Wertsch, 1996; Engestrom & Miettinen, 1999; Wertsch, del Rio, & Alvarez, 1995). Some have emphasized that these activities may be understood as joint work practices and that individuals participate in these practices as part of a community of others (i.e., a community of practice per Lave, 1996; Lave & Wenger, 1991; Rogoff, 1994; Rogoff, Baker-Sennett, Lacas, & Goldsmith, 1995; Wenger, 1998).

Within these communities of practice, various supports or "scaffolding" help learners deepen their engagement in particular work practices (Vygotsky, 1978). These supports include assistance from others more deeply engaged in or experienced with those practices (e.g., Derry et al., 2000; Tharp & Gallimore, 1988; Wenger, 1998). These forms of assistance move beyond generic calls for districts to send coaches or to deliver new information about educational improvement to schools. Rather, these forms of assistance involve relationships that make particular resources available to principals and teachers. As we elaborate next, these resources include brokering, new models of professional practice, valued identity structures that reinforce those models, dialogue-rich social opportunities, and tools that focus practitioners on particular "joint work." We suggest that a third party such as an intermediary organization may be particularly well suited to these activities. Many of these activities demand an ability to demonstrate challenging teaching and leadership practice that may be rare within district systems—or else instances of learning improvements would be more common than they seem to be. In addition, many of these activities are substantial areas of work that may lie beyond what district practitioners can typically add on or integrate into their own ongoing professional demands.

### *Brokering*

Wenger and others have emphasized that participants in assistance relationships enable learning in part when they operate as brokers or boundary spanners—individuals who move between communities of practice and their external environments (including other communities of practice). Brokers may bridge communities to new ideas and understandings that may advance their engagement in particular work practices. They also may buffer those communities from potentially unproductive ideas and understandings (Wenger, 1998). Brokering seems particularly productive when brokers do not simply pass knowledge resources across organizational boundaries but translate them into forms that the receiving community may be particularly likely to understand and use (Cobb & Bowers, 1999).

## *Modeling*

Participants in assistance relationships support engagement in new work practice by modeling or making available those who model forms of practice (e.g., school leadership, classroom teaching) that foster particular outcomes (e.g., high-quality teaching and learning; Brown & Campione, 1994; Tharp & Gallimore, 1998). By observing and systematically analyzing models, practitioners may develop a conceptualization of the new work practices prior to engaging in them—conceptualizations that theorists argue are essential to execution especially at deep levels of participation (Collins, Brown, & Holum, 2003). Such conceptualizations provide

an advanced organizer for the initial attempts to execute a complex skill, . . . an interpretive structure for making sense of the feedback, hints, and connections from the master during interactive coaching sessions, . . . and . . . an internalized guide for the period when the apprentice is engaged in relatively independent practice. (Collins et al., 2003, p. 2; see also Lave, 1996)

Furthermore, models sustain practitioners' engagement in particular promising endeavors by infusing those endeavors with value and increasing practitioners' confidence that they may be on a trajectory to deepen their engagement in those work practices (Brown & Duguid, 1991).

Particularly powerful models employ meta-cognitive strategies of bringing "thinking to the surface" and making it "visible" (Collins et al., 2003, p. 3; see also Lee, 2001)—for example, by engaging others in dialogue about the purposes and nature of the practices—so others know not just what participation in these practices entails but why they should participate in particular ways. Powerful modeling also involves a strengths-based approach in which the modeler helps others to identify their strengths and to build on those strengths to develop other competencies (Lee, 2001). These forms of modeling involve not a generic set of supports but assistance reasonably fine tuned to the developing capacity of all participants.

Some elaborate that particularly powerful forms of modeling are reciprocal (Tharp & Gallimore, 1988; Wenger, 1998)—that in helping others deepen their engagement in particular work practices, modelers also continually examine and transform their own participation in the process. In this view, assistance becomes a mutual learning relationship.

### *Establishing and Reinforcing Valued Identity Structures That Legitimize Peripheral Participation*

These structures include markers that indicate progressive degrees of participation such as the badge system in the Girl Scouts (Rogoff et al., 1995). Such identity structures help participants recognize that even those who are not yet participating fully in a particular work practice may nonetheless be on a trajectory toward deeper participation and as such they are valued members of the community (Wenger, 1998). Some call such participation novice or “peripheral” in part to signal that it is on the outside but somewhere within the range of stronger performance. They argue that individuals tend to deepen their engagement in various activities when they see themselves as valued participants in the activities and as people capable of deepening their engagement, regardless of their starting capacity.

### *Creating and Facilitating Dialogue-Rich Social Opportunities*

As previously noted, social engagement is fundamental to learning; the active construction of meaning unfolds not mainly within practitioners’ minds but as practitioners interact with one another (Weick, 1995). Through social interactions within communities of practice participants increase the individual and collective knowledge brought to bear on situations (Lave & Wenger, 1991; Wenger, 1998; see also Holland, Lachicotte, Skinner, & Cain, 1998). Through dialogue, participants have opportunities to challenge each others’ beliefs and interpretations of problems and events. Such dialogue can lead to new shared understandings and deeper engagement in particular activities than would otherwise be possible by individuals operating alone (Brown & Duguid, 1991). The models and identity structures just discussed may operate as resources for learning only provided community members have opportunities for social engagements with others through which they may observe those models in action (Wenger, 1998).

### *Developing and Continually Elaborating Tools*

Tools are “reifications,” the manifestations of ideas (Wenger, 1998) or, in simpler terms, the specific form that new ideas about work practices may take. Tools help deepen individuals’ engagement in particular work practices by “specify[ing] the parameters of acceptable conduct,” communicating messages about what individuals should and should not do. At the same time, “their meaning is not invariant but a product of negotiation with a community” (Brown & Duguid, 1991, p. 33). Accordingly, tools also operate as jumping-off points for practitioners to define new conceptions of acceptable conduct. These structures can serve as origins or “the kernel that provides the pretext for assembling” elements in the first place



(Weick, 1998, p. 546). As such, tools do not prescribe action but “trigger” negotiations among individuals about which actions to take toward meeting particular goals (Brown & Duguid, 1991). They may “be seen as liberating in their enabling function or limiting in that their historical uses may preclude new ways of thinking” (Smagorinsky, Cook, & Johnson, 2003, p. 1407).

*Conceptual tools* include “principles, frameworks, and ideas” (Grossman, Smagorinsky, & Valencia, 1999, p. 13). These tools are generally designed primarily to frame how people conceptualize particular problems or issues. *Practical tools* provide specific examples of “practices, strategies, and resources” that have “local and immediate utility” (Grossman et al., 1999, pp. 13–14). Conceptual tools aim to shape participation across multiple activity settings whereas practical tools are generally constructed around particular types of activity settings.

### *Focusing Engagement in “Joint Work”*

Assistance relationships appear to foster learning when they focus participants on “joint work” (also called a “joint enterprise” or “authentic situation”; Rogoff, 1994; Rogoff et al., 1995; Wenger, 1998; see also Brown, Collins, & Duguid, 1989). Joint work includes activities that participants value and that promise to help deepen their engagement in particular forms of work practice. Accordingly, the concept of joint work serves in part to reinforce the reciprocal nature of the assistance relationships by emphasizing participants’ engagement in activities that all parties find meaningful. People in assistance relationships support engagement in joint work by providing opportunities for others to co-construct the meaning of particular challenges and the potential fit of strategies to address those challenges (Wenger, 1998).

## RESEARCH DESIGN AND METHODS

These learning theory concepts grounded our analysis of the IFL’s engagement in assistance relationships with eight urban districts. We used a retrospective, cross-sectional, and largely qualitative case study design. Retrospective data on IFL history allowed us to examine the IFL’s theory of action about how to support districtwide learning improvements from its inception in 1995 to the start of our real-time data collection. The cross-sectional component helped us interrogate how the IFL operated in practice as that practice unfolded. A qualitative focus seemed particularly appropriate given the situated nature of the IFL’s work and the importance of capturing rich accounts of their work processes and how participants made sense of them as described next.

### *The IFL Case*

The IFL provided a rich and particularly appropriate case for this inquiry. Since 1995, the IFL has worked with several districts with a main goal of strengthening learning for all students districtwide (Glennan & Resnick, 2004). Accordingly, the IFL stands apart from some other intermediary organizations in its districtwide change focus. Assisting district performance was an explicit, core dimension of their district partnership strategy. The founder and director of the IFL, Lauren Resnick, and other IFL staff reported that they aimed to work with key central office staff and principals to build their own capacity to assist teachers and others with learning improvement efforts. As Resnick explained,

We're trying to build a professional development system that will train teachers eventually but our work will tend virtually always to be with those in the district who do the teacher training . . . . So it's our job to build that district capacity.

At the time of our study, the IFL employed approximately 25 full-time staff people called "fellows," whom it deployed to districts to participate directly in these assistance relationships on behalf of the IFL. In interviews over time, these fellows invariably referred to their intended roles as assisting with districtwide student learning improvements by helping build local capacity for such outcomes.

Since its inception, the IFL has received core operational support from the University of Pittsburgh as well as from several national foundations. The IFL's partner districts provide an ongoing source of revenue by paying for certain IFL services. The IFL's relatively long-term success in garnering diversified funding suggested it might demonstrate activities of a relatively high-functioning intermediary not impeded by the predictable pitfalls of an organization in start-up mode or regularly threatened by significant budget constraints.

Furthermore, the IFL seemed to offer a case of an arguably "successful" intermediary organization. The RAND Corporation and others across multiple studies have associated the IFL's efforts with particular outcomes that suggest IFL's assistance may be helping districts advance along a trajectory of deeper engagement in supporting learning improvements (Marsh et al., 2005). These outcomes include shifts in central office administrators' thinking about and engagement in supports to school principals, the development of formal district policies that aim to enable this thinking and participation, indicators that school principals are working more closely with their teachers to enrich their practice, principals' increased skills and knowledge in specific content areas and pedagogy generally, and some very modest improvements in teachers' practice.

However, prior research on the IFL is limited in that it did not attempt to link the IFL's work with actual changes in student performance. We acknowledge this limitation here and address it in the framing of our findings. Specifically, we do

not claim that the IFL's activities have directly caused improvements in student learning. Rather, we emphasize that prior reports have associated the IFL's work with some changes in how school principals and central office administrators think about and engage in their work in ways that theory and experience suggest matter to learning improvements. We aim here to elaborate the features of the IFL-district relationships associated with those changes in professional work practice.

This approach seems particularly appropriate given the contingent nature of the IFL's work and the nascent stage of research on how to tie such work to student learning outcomes. As Resnick articulated,

We don't know that we can trace everything that's going on there [a partner district] to us, but we don't intend to anyway. . . . Unlike the whole school model that's trying to say "Do it this way." . . . [We] work with the districts . . . to help them choose and coordinate all the different kinds of things they're doing [rather than create a new intervention for which we can unambiguously take credit]. . . . We can't be the direct solver of the problem. But once we see it clearly . . . if the time is right. . . . Then what we hope to do is not to be the designer of the solution but to be in that conversation so that the solution comes out instructionally as powerful as possible.

Various comments by other IFL staff and district practitioners confirm such difficulties in tracing specific outcomes to IFL actions. As one IFL staff member explained,

There's lots of things that I could point to and say, yeah . . . that was something that the IFL pushed. . . . But . . . if you asked [a district central office administrator] where she got those ideas from . . . I've seen her leave [IFL] retreats and sessions saying, I learned nothing today and then six months later I see her enacting some of the things that came up in the meeting. Is she conscious that there might be a link between the two? I don't know.

Given these considerations, we chose an intermediary organization that other research associates with helping shift district work practices in ways that may be associated with student learning improvements. We aim to elaborate what this organization does and, where possible, to link those activities with reports from various reform participants about relationships between activities and outcomes. Such examinations of intermediary work practice can provide important anchors for future outcome studies as we discuss in our concluding section.

### *Data Collection*

We draw on data collected between 2001 and 2006. Data from 2001 through 2005 come from two mixed-methods studies of the IFL's relationships with eight

urban districts conducted by one of the paper authors and colleagues at the RAND Corporation (Marsh, Kerr, Ikemoto, & Darilek, 2006; Marsh et al., 2005). These investigations surfaced a sizeable dataset, but reports on these investigations to date have been limited to basic descriptions of IFL activities (e.g., frequency of meetings with staff) and not grounded in an explicit theoretical framework. Between 2005 and 2006, both of the authors of this article conducted additional data collection activities in one of the original eight districts and in another district that initiated a partnership with the IFL within the past two years. We framed this second wave of data collection centrally around the conceptual framework highlighted previously.

*Interviews.* We reviewed notes and transcripts from interviews with 251 respondents conducted between 2001 and 2005: 80 district and community leaders, 73 principals, 30 assistant principals, 50 school coaches, and 18 IFL leaders and staff. Between 2005 and 2006 we conducted additional interviews with 11 IFL staff members—including four who had not been interviewed during previous data collection—and 14 district leaders, among them nine who had not been interviewed during previous data collection. The district interviewees were superintendents, chief academic officers, supervisors of principals, and other leaders directly involved with the districts' IFL partnership. All interviews focused on the nature of IFL's district work including the rationale for particular approaches and how participants made sense of how the work was unfolding in real time.

*Observations.* We reviewed notes from approximately 200 hours of formal meetings that occurred between IFL staff and district practitioners and among IFL staff between 2001 and 2005 (e.g., annual retreats for IFL partner districts, meetings on site with district principals, and IFL staff meetings). We conducted an additional 32 hours of meeting observations between 2005 and 2006. These observations focused on the extent to which IFL activities reflected or departed from the concepts elaborated in our conceptual framework.

*Documents/artifacts.* We reviewed more than 150 documents that captured the evolution of the IFL's district partnerships over time including multiple versions of IFL-authored descriptions of their work, IFL's tools, records of the IFL's district plans, and artifacts from IFL trainings. We also included reported and unreported descriptive analyses of IFL work written by RAND researchers over the course of their research.

*Focus groups.* We reviewed transcripts from teacher focus groups that included 118 teachers across three districts. We included in our analysis for this



article those portions of the focus group conversations that addressed teachers' experiences with IFL staff and activities.

### *Data Analysis*

We used NUD\*IST (QSR6) software to code our data in several phases. First, we used low-inference categories to sort through basic dimensions of the IFL's district partnerships including the IFL's intended relationships with districts, instances of the IFL–district relationships in practice, outcomes that seemed associated with these relationships, and conditions that seemed to help or hinder those relationships either by respondents' direct reports or our observations. Second, we recoded our data using higher inference concepts from our conceptual framework including examples of brokering, modeling, providing dialogue-rich social opportunities, tool development, and engagement in joint work. We catalogued as “other” any data that seemed to capture important dimensions of the IFL's district relationships that did not fit obviously into these categories. We also asked IFL respondents and a RAND researcher who led their IFL research to review our draft report carefully and highlight consistencies and inconsistencies with their interpretations of events. We used such respondent reviews as an additional check on construct validity and the overall reliability of our analyses.

We acknowledge that interviews served as a primary source of data for this analysis and that such self-report data can be a poor substitute for extended observations of actual practice. For example, our colleagues noticed that district central office leaders tend to be more positive about the IFL's work than school-level leaders. Our own experience with various implementation studies suggests that respondents may tend to report significant frustrations when engaged in challenging work such as that supported by the IFL and to magnify their difficulties beyond what might be documented by more dispassionate observers over time. We dealt with these potential biases in self-report data by triangulating reports from different types of respondents—school-level staff, district-level staff, and IFL staff. Whenever possible we corroborated self-reports with data from observations and documents. Throughout our report of findings we indicate whether we derived a particular claim from self-report or another data source to help readers judge the bases of our claims.

## FINDINGS

Analyzing and interpreting the day-to-day work of an organization with complex goals poses significant challenges. The majority of our data confirm that IFL staff tended to work in partnership with district practitioners in ways consistent with the activities in our conceptual framework. However, we also found examples in

which IFL staff deviated from those activities intentionally, because of limited internal capacity for engaging in them, or because of other factors. In fact, one of our main findings is that the work of learning-support intermediary organizations is inherently challenging and dependent on the participation and capacity of others well beyond their control. Such work may wax and wane along various dimensions of success over time. Where possible in the upcoming sections, we note significant counterexamples to the predominant patterns. But, given the complexity of these processes, space limitations, and the nascent stage of research in this area, we focus our discussion on elaborating predominant features of the IFL's assistance relationships.

Overall, we found that the IFL–district assistance relationships seemed to involve brokering, modeling, particular social opportunities, and the development and use of tools, all centered on joint work or particular problems of practice identified and elaborated by both IFL staff and district practitioners. We did not find sufficient evidence that the IFL created valued identity structures. In certain modest respects, the IFL's activities also extended beyond those in our conceptual framework. For example, the social opportunities enabled by the IFL–district assistance relationships seemed to reflect those anticipated by theory; however, characteristic of these opportunities also were IFL efforts to lessen the risk district practitioners may have associated with engaging particular new work practices. Across all these activities, we found evidence that the IFL regularly revisited and occasionally revised their approach to engaging in those activities within some foundational parameters as district practitioners' and IFL staff both deepened their engagement in particular work practices and as district conditions shifted. To reflect and reinforce this cross-cutting feature of the relationships, we call them "adaptive assistance relationships." In the following subsections we elaborate these dimensions of the IFL's adaptive assistance relationships with its partner districts.

### *Brokering*

Interviews and observations indicated that all IFL fellows at least occasionally operated in ways consistent with what theory refers to as "brokering"—linking district practitioners to a variety of new ideas (and people with ideas) about how to strengthen student learning. As one IFL staff member elaborated,

That's [brokering is] . . . true for . . . most of us. You know, you get into these meetings [with district staff] and . . . you become a purveyor of . . . all the ideas that are out there. We work here [at the University of Pittsburgh] in such a way that we hear what others are doing and we know what other districts are doing. So when we're in meetings with these folks we can say, well, you might want to think about this or you might want to go there. But I don't think that's necessarily just a role that I have, I think it's one that probably happens with most of the [fellows].

Most of the examples of IFL fellows operating as brokers that we surfaced in our data involve IFL fellows linking district practitioners to research on how people learn. By many IFL staff members' accounts, Lauren Resnick and others at the University of Pittsburgh first developed the IFL in the mid 1990s to bridge district leaders involved with the New Standards Project to research on learning that both district practitioners and early IFL staff believed would help them realize their early standards (Glennan & Resnick, 2004; Honig & Ikemoto, 2006). In interviews, virtually all the IFL fellows cited multiple instances in which they consulted various research databases for such resources. For example, one fellow described the following strategies she or he used to prepare for professional development sessions with his or her districts:

I'll speak for myself, but I know my colleagues do this. When I'm putting together a [professional development] session, I spend a lot of time on the internet, googling different authors or different ideas or like different concepts to see what pops up. . . . At the [annual convening of member districts] . . . our group was doing the . . . professional development. So one of the things that I did was spend about, I don't know, two or three days hanging out here at home on the computer, looking up all kinds of things, in terms of coherent professional development. And out of that surfaced those characteristics that we use, but also out of that surfaced like three or four articles that were right on target for that particular topic and very informative and that now people are using. And that's true for all of us [fellows], you know.

IFL fellows also linked their partner districts to researchers themselves. For example, researchers presented at all of the IFL's annual meetings for its partner districts between 2000 and 2005. In one instance a nationally prominent researcher shared her work on assessment during a 2005 retreat for the IFL district partners. Our observations and document reviews suggested that IFL staff members also linked district practitioners to new ideas from research by engaging researchers in developing materials specifically for the IFL fellows to use in their work with districts. (We elaborate on the process of tool development in the subsection on "tools" next.) For example, during our data collection period, IFL fellows and other staff worked with a school leadership researcher to develop a principal evaluation rubric and to pilot that rubric in one of their districts. Occasionally, the IFL hired researchers as part-time staff to provide ongoing consultation. For instance, in recent years, several researchers have joined an IFL work team to share their research knowledge regarding secondary school content areas. As one IFL staff person confirmed,

We really do try to bring in researchers . . . to talk to us, to talk to the district people. And we want to be informed not just by reading the articles, but then having them [researchers] look at our materials and say, "Does this fit within the realm of what you believe you mean when you say this?" . . . So when we did a lot of work on [a

particular set of materials for districts] we brought in researchers who do the research on talk [dialogue that promotes learning]. . . . And it was a marriage between . . . research and practitioners going back and forth about what it means to have talk not only from the research-base, but what we've learned in classrooms.

IFL staff also bridged district practitioners with researchers by inviting researchers to conduct research in their districts. For example, the IFL asked the RAND Corporation to conduct several studies including analyses of the IFL's participation in supporting district reforms in three of its member district (and from which we draw some of the data presented here). At the end of each study, IFL staff invited RAND researchers to share their findings with all IFL staff, and in each instance they devoted almost an entire day to a discussion about the findings and how fellows might use those finding to improve their work with partner districts.

IFL fellows also frequently bridged their partner districts to each other and lessons learned about promising teaching and learning improvement efforts in other districts. For example, our interviews with and observations of district practitioners confirmed that across our focal IFL districts, fellows routinely created opportunities for district central office administrators and school principals to observe practitioners in other districts engaging in Learning Walks—a strategy IFL staff derived from their partnership with then New York City Community School District #2 to help educational leaders and teachers observe, analyze, and improve teachers' practice and supports for that practice (Institute for Learning, 2003). In a comment typical of our data from district practitioners, one district leader explained,

For me, personally, as a new [district leader], [the IFL partnership] was an excellent way for me to get mentoring and coaching and to really hook into a professional community very quickly, because I don't know that I would have known how to go out and find other[s]—and I couldn't—but I very quickly got in with a group of other [district leaders] in cities with similar kinds of schools, and conditions and situations that [my district] had.

According to some IFL fellows, linking district practitioners to one another was one of their most important bridging activities in part because they believed many practitioners in their particular districts were on the cutting edge of knowledge about how to strengthen teaching and learning districtwide. According to one fellow,

I think we're learning so much. . . . What some of these successful districts are both taking from our work but also what they're turning them into on their own is very useful because they're ahead of where the research is. . . . So we want to make some case studies and . . . give the world some of the examples of what's going on in these districts that would inform the rest of the world.



Across these and other examples of the IFL's brokering activities, IFL staff did not simply pass new ideas to district practitioners but grappled with where to look for particular knowledge resources, what they were finding, and whether and how to share it with district practitioners in ways consistent with what theorists refer to as "translation". For example, one IFL staff member described her brokering efforts as fundamentally about understanding the meaning of various new information and how to frame it for district practitioners,

I take notes on the article. I then think to myself how it's useful. I then try to do some—I guess probably what I do research-wise is do some triangulation, look at some other articles by other researchers, ones that I recognize, ones that I know. I might look at those and then I try to see how what one says coordinates with what others say. And then after I do that then I might sort of put it together. I might frame it, put it together. Then I would send it out. In the case of [developing one set of materials], I . . . sent . . . out the copies of the articles [to other fellows. I asked them,] "How does this sound?" So in effect, I used my colleagues to vet the material. . . . Then we might sit and talk about it. Then after that, then start to think about, "Okay, how will we use this?" . . . So then we would go back and forth.

In self-reports of their work between 2001 and 2006, IFL fellows routinely indicated that a key dimension of their work involved consultations with each other about how to translate research resources into forms that might help their district partners integrate the research into their own work practice. As one fellow described this process,

There's a group of us, probably I guess that practically the majority of us, when we come across something that we think has applicability in other settings or it's some research that we think is really interesting to think about using, we usually share it and, you know, send it out for people to see. And then what usually happens . . . next is that people will say, "That's really good. Did you use it?" And then, of course, you know you say, "Yeah, here's the task sheet." Or, "Here's the protocol." Or, "Here's how I used it and here's how I ramped up to it and then here's what I did with it and then here's what I ask people to do with it."

We also found that IFL staff members' efforts to link district practitioners with specific research communities reflected particular biases. For example, IFL staff appeared to readily link districts to research on adult and student learning and the importance of trust and tools to learning. Interviews and observations suggested that staff tended to favor this research in part because they had personal connections to individual researchers in these areas. When we probed about their engagement with other research that seemed relevant to the IFL's work (including research on implementation, change, and leadership), IFL staff generally reported that they infrequently searched for or used research in those areas. IFL staff's generally

limited familiarity with this research may have curbed their searches within these research communities for research information that could have grounded their efforts.

A few district leaders suggested in interviews that they picked up on these biases and that they viewed the IFL's relatively weak knowledge of research beyond learning research as a limitation on their brokering roles and work overall. For example, one district leader reported that what he or she perceived as the IFL's narrow focus on instructional leadership resulted in the IFL's principal leadership training leaving out important aspects of school leadership. In this person's words,

I think it might behoove them to look a little bit more at [research on] the role of the principal and to not let ideology color their view of what a principal should be doing. So I think they need to do a little bit more work on how the principal can work effectively as an instructional leader while having organizational responsibilities.

### *Modeling*

Data across our entire study likewise surfaced multiple examples of IFL fellows using modeling strategies to engage district practitioners in new work practices. For example, during visits to districts, IFL staff routinely modeled or linked district practitioners to others who modeled how to conduct LearningWalks. In one typical instance, an IFL fellow demonstrated how to look for evidence of what they considered high-quality teaching and learning while observing classrooms as part of the IFL's LearningWalk protocol. A central office leader in that district reflected that part of why the IFL has been so helpful with supporting new leadership practice is because IFL fellows have been present on site in his or her district demonstrating these new work practices, observing district practitioners engaging in them, and helping district practitioners make sense of the new work.

Our observations captured multiple instances in which, as part of their modeling efforts, IFL fellows made thinking explicit in ways consistent with the meta-cognitive activities described in our conceptual framework. For example, when asked to identify what she considered the most important features of her work with district practitioners, one fellow highlighted, "The idea of making thinking visible . . . the meta-level of stepping back and reflecting on what supported your learning, and what the implications of that are for what you're going to do when you try to support someone else's learning."

All the IFL fellows interviewed not only led professional development sessions but, in the course of the sessions, they regularly and explicitly labeled strategies they were using. For example, in one session with district staff that we observed, the IFL fellow serving as facilitator led participants through establishing norms to guide their conversation as a group. She also explained to participants that

she was trying to help them establish such norms with the hope that doing so would facilitate the kinds of direct, honest, and sometimes difficult dialogue that reflecting on professional practice required. When asked in an interview to explain the rationale for this approach to her dialogue with district practitioners, she explained that she frequently reflected her strategies for running meetings back to participating school principals and central office administrators. She thought that such a meta-cognitive strategy would help principals identify these strategies and would increase the likelihood that they would use similar strategies in their work with their own staff in other settings than if she simply demonstrated the activities without making her thinking explicit.

We also found some evidence that reciprocity characterized the IFL's modeling. As one IFL staff person explained, through their relationships with districts they too interrogated and deepened their own participation in particular work practices,

[We were] . . . asking them to read the [research] articles, to think about what's going on in the school, and then it was their questions and their comments back that made us kind of tweak and refine the list [of characteristics of high quality professional development] that we ended up with. So it was actually watching it play out in schools and what's possible and what's not possible, what really makes sense to teachers and what makes sense to principals. So I got as much from it [the relationship] as probably they got out of it. . . . That's the two-wayness.

Another IFL fellow corroborated, "We learn from the people we work with every-day."

IFL staff also solicited feedback from their partner districts during and after all of the formal IFL-district meetings observed in our period of study. Distributing feedback forms at meetings is not a particularly unusual, uncommon, or necessarily significant activity. However, the IFL's feedback efforts seemed to indicate reciprocity in their relationships because IFL staff routinely used the feedback to interrogate their own participation in the assistance relationships and how they might improve on it. According to one IFL fellow, the IFL's use of feedback reflected how they relied on their relationships with practitioners to help them enhance their own work:

I think that we all really rely on the feedback forms—the field, the coaches, who are trying to use our ideas and our materials and tools . . . —to keep us honest and make them [our sessions] really useful. So, I would say . . . we definitely rely on the practitioners, the practition of coaches and coach coordinators that we work with.

The reciprocal nature of these relationships seemed fundamental to district practitioners' willingness to participate in these relationships over time in ways that seemed to matter to deepening their engagement in particular work practices.

As one district central office administrator said in reflecting over a six year partnership with the IFL,

I think some of things [this district] has done with [the IFL] has shown up in their [the IFL's] work. I know our literacy coaches will point out that certain things the IFL now says are based upon their experiences within the district, or . . . a push-back they got from the district. . . . The IFL tries to tailor—tries to make some accommodation to what the system already does, or they may even rethink . . . some of the truths that they held based upon what they see happening. We see it as more of a dynamic body of knowledge that, yes, they present knowledge to us in a structured form that we may not have thought about, but they also are willing to adjust that to the on-the-ground reality sometimes.

Our data also supported the importance of particular forms of modeling to district improvement efforts by negative example. In two districts, central office administrators in particular reported that the IFL had a limited impact on their formal policies and how they participated in teaching and learning improvements. These central office administrators tended to describe the IFL's work in their districts as overly theoretical and lacking concrete forms of support. Some individuals talked about wanting the IFL to “connect the dots”—provide more concrete examples and direct follow-up with IFL fellows. For example, one of these central office administrators reported,

The only thing I guess I would like more of is . . . how to get things accomplished. I think lots of times the Institute, and that's probably the way it's designed . . . gives the questions, facilitates discussion, but doesn't really give you the answers. Sometimes you'd just like to have more answers or more best practices from this other district, more real examples of how to make things happen, rather than just discussions.

### *Social Opportunities*

As part of their assistance relationships, IFL staff engaged district practitioners in a countless number of formal and informal dialogue-rich social processes. The extent of the formal social processes seemed to vary by the IFL's contract with partner districts—specifically the number of on-site technical assistance days they had negotiated with each district and accordingly how many opportunities they had to convene groups of district practitioners in meetings. Regardless of the specific contractual terms, IFL fellows typically held monthly day-long professional development sessions for particular stakeholder groups within each district, such as central office administrators, principals, and/or school coaches. Some conferences or professional meetings may be characterized by formal presentations or the transmission of information to attendees. By contrast, the IFL's day-long



professional development sessions fundamentally involved dialogue among district practitioners, as well as between district practitioners and staff. For example, our reviews of formal meeting agendas suggested that significant portions of these meetings were dedicated to analyzing research articles in small groups, examining and discussing videos of classroom instruction, and sharing their reflections on how the new information related to their own professional practice.

IFL fellows typically deliberately structured such dialogue to provide district practitioners with opportunities to socially construct the relevance of particular ideas or forms of work practice to their own ongoing practice. As one district leader explained, these opportunities to grapple with the meaning of new ideas in light of their own practice had a much greater impact on his or her practice than exposure to the same ideas absent opportunities for social construction:

In my work in graduate school, I had some experience with WalkThroughs, which is the early iteration of LearningWalks. And I'd had my requisite courses in psychology of learning and all that stuff, but I hadn't really had the opportunity to practice it. So I've learned enormously from the work [with the IFL].

This district leader elaborated that “the opportunity to practice it” included written and oral guiding questions from IFL staff that engaged him or her and his or her colleagues in making sense of why to conduct a LearningWalk, what were the basic parameters of this activity, and how this leader might execute it in his or her own context. Another district leader corroborated that a key dimension of their relationship with the IFL involved IFL fellows assisting them in making sense of the implications of new ideas for their ongoing practice,

[The IFL has been] helping us to translate that [research] into good practice and then really applying it to our own situation. So translating the research to practice across the country and then adapting it to our own situation, I think, was the perfect flow.

Another district leader reported in an interview that the IFL routinely carved out significant portions of their meetings for him or her and his or her staff to learn about new ideas and to jointly discuss how new ideas mattered to their work. This leader reported that these opportunities were a fundamental aspect of his or her work with the IFL and essential to the work permeating ongoing professional practice.

I think that there's a tremendous need for the work to become personalized for a district and very connected with the work that the district is doing in order for it to be meaningful so that it doesn't become a layer. When it's visualized as just a layer and not embedded in the work that you're doing, then it doesn't have meaning anymore.

These social opportunities also seemed to challenge district practitioners' beliefs and prior knowledge about how to support teaching and learning districtwide. For example, a longtime IFL staff member commented that she had observed IFL staff frequently talk about the importance of creating "a sense of disequilibrium" among the practitioners participating in IFL sessions. One fellow went so far as to report that she viewed her work as a failure if she did not create that sense of disequilibrium. She explained that such instances are

when people are really learning something. . . . [As one fellow talks about it] The way that she knows that that's going to happen is when she gets somebody saying, "Oh my God, wait a minute. I never realized that I was doing this," or, "I didn't realize I wasn't doing this." And so she's come up with ways to get people to reach that point of disequilibrium. Sometimes it's a lesson that will challenge them to think about something that is uncomfortable for them or that stretches them a little. Or she'll have them make videos of their own practice and bring them in. And [she's] not the only one who does this, by the way. This is something that all of the [IFL teams] do [with practitioners].

In interviews, district practitioners too characterized their discussions with IFL fellows as challenging their thinking. As one district leader put it, "The system has never been ready to do any of the things they want us to do, so it's a constant push. They're a pushy partner, which is exactly what the system needs."

Our observational data, in particular, suggested that the IFL staff worked to construct these social opportunities in ways that limited (or that they intended to limit) risks district practitioners may have associated with examining and critiquing their own practice. For example, as previously noted, our reviews of meeting materials from more than 30 IFL professional development sessions revealed that IFL staff often invariably kicked these meetings off with a discussion of norms (often referred to in meetings as norms of "successful professional learning communities") that the group would use to guide their work in and beyond the session. To reinforce these norms, the IFL fellows that led these conversations often modeled what they called "talk moves"—how to disagree with another speaker in a respectful manner. Observation notes also indicated that IFL staff created norms of safety around specific tasks. For example, in one instance when school principals were observing a video of a guided reading lesson in one classroom, observers recorded multiple instances in which the fellows explicitly discouraged participants from evaluating the teacher and instead focused their attention on the extent to which the video included evidence of powerful teaching and learning.

### *Developing and Supporting the Use of Tools*

The IFL also devoted significant staff and fiscal resources toward the development of tools that they used to ground their district assistance relationships. (For an elaborated analysis of the IFL's efforts to develop and support the use of tools; see Ikemoto & Honig, in press.) Consistent with theory, IFL staff in interviews distinguished tools from other materials in that tools carried ideas they believed practitioners would value and use and in that they usually promoted particular activities to engage district practitioners in those ideas. For example, when asked to explain what makes a particular IFL document or set of materials a tool, one IFL staff person explained,

The tool has to carry the theory as well as the action because you just can't tell people to have an invested learning community; you have to put them in one. And . . . that was based on some research on tools, on what we do to develop good tools. . . . We had to build the tools that produced the action rather than tell people to have the action.

According to another, "a tool should make people not so much believe in [the importance of a given activity] but it should actualize it. . . . Then they can step back and say, "Oh that's what you meant by it." So in the tool's very essence of being used, it should make a person live out [particular activities]."

The tools that IFL staff created during our period of study could be classified as either primarily conceptual or practical. Conceptual tools were mainly text-based statements designed mainly to present particular ideas. For example, the "Principles of Learning" were essentially nine statements about the characteristics of environments that promote rigorous instruction. Practical tools had conceptual dimensions but emphasized action rather than ideas as the main avenue for helping district practitioners incorporate particular ideas into their practice. For example, the LearningWalk tool rested on a set of ideas about how to engage teachers in critical examination of their own practice but took the form of a series of guided activities for principals and others to use to foster such examination.

IFL staff developed two different kinds of conceptual and practical tools. One type, that we labeled "local tools," included situation-specific tools usually created by individual fellows to address a particular challenge within a given district. For example, in one district the IFL-district partnership focused in part on strengthening principals' support for reading instruction. Several fellows working in that district developed a tool that engaged school principals in research on the role of fluency in learning to read and then examined the extent to which ideas from this research were reflected in their own local state standards and district literacy curriculum. Other tools, that we called "organizational" tools, were developed for use in multiple districts. For example, fellows and full-time staff across the IFL

contributed to the development of the LearningWalk tool, and IFL fellows reported that they had used this tool in almost all of their partner districts as one resource for shaping district practitioners' work practice. In some instances, IFL fellows over time developed particular local tools into organizational tools. For example, several fellows created a number of practical tools for engaging principals in the Principles of Learning. Although these fellows consulted with one another, they largely crafted the activities for the specific context and purpose of their assigned districts. These fellows later came together to create an organizational Instructional Leadership Program tool that integrated their local work into a resource for use across IFL districts.

A recurrent theme in our data was that the IFL drew on particular resources as the basis for the development of all their organizational tools and most of their local tools—namely, ideas from both research and their own as well as their district partners' experience. This use of both research and practice knowledge seemed a particular hallmark of the IFL's approach to tool development and, by some self-reports and our own observations, essential to district practitioners' sustained use of certain tools over time. For example, early IFL staff began developing the Principles of Learning by reviewing certain research on how people learn and then distilling that research into selected key dimensions of powerful teaching and learning environments. However, the Principles of Learning now in use at the IFL developed over the course of at least four years through formal and informal conversations with a core group of educators and IFL staff during which IFL staff and district practitioners grappled with the value of particular research-based ideas and how to word particular complex concepts in ways that might resonate with practitioners (Resnick, Hall, & Fellows of the Institute for Learning, 2001). Other tools such as the LearningWalk began with practice of district central office administrators and school principals—in this case, those in New York City Community School District #2. In the early 1990s, then superintendent Tony Alvarado began supporting his staff in observing each others' practice and engaging in challenging conversations around their observations and student work. District and IFL leaders observed that these activities seemed to be having a demonstrable impact on teachers' and principals' work practice and that the activities, albeit unintentionally, reflected research on how people learn. Over a series of years, with the help of researchers from the Learning Research Development Center, which housed the IFL at the University of Pittsburgh, IFL created the LearningWalk tool that incorporated lessons from District #2 practice, as well as research on how people learn, and eventually other research on professional consultations and trust.

The development of tools alone did not ensure their use in productive ways. Our data surfaced various instances of IFL fellows expressing concern that district practitioners occasionally misappropriated their tools. For example, one fellow explained how the IFL had attempted to make one of the Principles of Learning,



Clear Expectations, concrete and practical for practitioners by encouraging administrators to look for whether teachers were posting objectives and scoring rubrics on classroom walls during their LearningWalks. However, principals in several districts used this LearningWalk protocol as a checklist in ways that emphasized the superficial features of Clear Expectations and failed to reflect the underlying idea that teachers should be making sure that their students were developing solid understandings of what they were expected to know and be able to do. Some IFL fellows, therefore, feared that practical tools might undermine the value of the ideas that the IFL was attempting to support. As we elaborate in another report of the IFL's work, the provision of the other forms of assistance, including modeling and social opportunities, seemed to significantly mitigate against misappropriation of tools (Ikemoto & Honig, in press).

### *Focusing on Engagement in Joint Work*

We found countless examples of IFL–district relationships focused on specific problems of practice, or what theory refers to as joint work. In some instances, the joint work involved challenges districts faced in improving teaching and learning. For example, one IFL–district partnership centered on how to improve the then disappointing implementation of a particular literacy initiative that the district had chosen to institute districtwide. Joint work also included district exemplars. The “work” in the latter cases became how to understand the conditions under which particular district activities seemed to promote positive results and how to use lessons learned from those exemplars to inform research and practice. For example, IFL staff reported that in the late 1990s they noticed that Tony Alvarado, superintendent in New York City Community School District #2, seemed to support teaching and learning successfully in part by reshaping the relationship between the school district central office and school principals. In the latter case, the joint work for the IFL, District #2 leaders, and Learning Research Development Center researchers became to uncover which activities and conditions seemed to contribute to the apparently positive results.

District practitioners typically prompted the forms of joint work that grounded the IFL–district assistance relationships, but the joint work generally took shape through ongoing deliberations between IFL staff and district practitioners as both parties grappled with the root causes of particular district challenges and how they could be addressed through the IFL–district partnership. For example, we observed two annual IFL staff meetings (in June 2003 and June 2004) in which

---

Teachers set Clear Expectations when they provide “clear standards of achievement and measures of students’ progress toward those standards that offer real incentives for students to work hard and succeed. Descriptive criteria and models that meet the standards are displayed in the schools, and the students refer to these displays to help them analyze and discuss their work” (Resnick et al., 2001).

IFL fellows described individual district work plans and how they were developed. According to the fellows' reports, the vast majority of the work plans came together through ongoing conversations between IFL staff and district practitioners:

In each district [how we work is] slightly different because . . . as you can tell by the resident fellows development of plans, it is the plans that we develop with the district that we're essentially responsible for delivering. . . . So it is co-constructed. I mean we clearly have in that construction an idea of what we think it is a district has to do that year to get someplace, but we also have to hear from them and have them agree that that's where they want to go. One of the things I think we do that's slightly different than most vendors . . . is that we customize. We are really attentive to the specific needs, realities, and circumstances of the districts we're in. And we're accommodating to . . . where they are and where they have to go. I mean, we differentiate the needs of our work in terms of districts as well as within the districts.

However, our analysis suggests that the extent of the co-construction of joint work may have been uneven during our study period. For example, some IFL fellows reported in interviews that district practitioners occasionally appeared reluctant to participate in the co-construction process—sometimes going so far as to call on the IFL to step into a situation and direct their work—and that the IFL–district plans developed under such circumstances reflected more of the IFL's ideas than district or joint ideas.

Some IFL staff expressed concerns that district practitioners occasionally demanded a didactic approach to help with their learning improvement efforts but that such direct guidance threatened to undermine opportunities for practitioners to grapple with ideas in ways that promised to deepen their understanding of those ideas. IFL staff generally reported that they managed this dilemma by “telling” practitioners explicitly how they should use ideas from IFL trainings early in their partnership and gradually lessening their reliance on this approach over time so that practitioners would take more responsibility for determining whether and how to apply knowledge to their practice. Accordingly, we view such early periods of direction as points on a trajectory of assistance toward joint work—strategies to help district practitioners engage in the co-construction of joint work over time. One district practitioner supported this interpretation that such directive periods were occasionally necessary to enable their deeper engagement in IFL activities over time:

I think in the beginning they showed up and said, “Here are the Principles of Learning. Here's a LearningWalk. Here's how to do it.” But now it's not that way anymore. I mean, that's because we didn't know anything. We needed that [in the beginning].

Another district practitioner corroborated.

In the beginning . . . it was the IFL who would tell us, “This is what needs doing and we’re going to come and this is what we’re giving you.” . . . And it’s almost like we’ve turned around [a corner on these mainly directive relationships] and said, “No, no, no, wait a minute. We know what you do. We know what you have.” And now we’re stopping and [we’re] saying, “These are the pieces that we still need in our district.” . . . They’ve just enabled us to come almost full circle, so I think we’ve grown as partners.

In other districts, excessive work demands may have truncated negotiations around joint work. For example, IFL staff reported in interviews especially in the last several years of our data collection that district practitioners presented them with more forms of joint work than they could realistically engage. These excessive work demands were so visible that even some district leaders reported to us that they would have liked to have involved the IFL in more aspects of their districtwide reform efforts but that the IFL staff seemed spread too thin to take on more responsibilities.

Although IFL staff reportedly wanted to jointly co-construct the work with each district, it also did not want—as one staff member put it—to “reinvent the wheel” in every district. Perhaps as a result, overtime, the IFL also tended to favor forms of joint work that seemed to fit well with tools they already had developed. For example, in the three districts from which we had particularly intensive interview data over time—and which were three of the IFL’s longest standing partnerships—the formal agreement about IFL–district assistance relationships emphasized four strategic areas: instructional leadership, coaching, curriculum specification, and the use of data in decision making. However, in practice, IFL staff spent most of their time and other resources on one of these areas, instructional leadership—specifically in working with district leaders to support principals in shifting from mainly managerial to primarily instructional support roles. At the time, the IFL had spent particularly significant resources on the development of their expertise in instructional leadership (including hiring staff with extensive experience in this area; see also, Marsh et al., 2005).

### *Adaptive Assistance*

We found substantial evidence that within the broad categories of activity previously elaborated, IFL staff often revisited, elaborated, and revised how they engaged in those activities over time as district and IFL capacity for particular work practices deepened, as other local conditions shifted (e.g., changes in district priorities), and as IFL staff deepened their understanding of different ways to realize those activities in their particular district contexts. In the words of one

IFL fellow, “we are continually changing how we do business.” This pattern was so prominent that we call the IFL–district assistance relationships “adaptive assistance” to capture this cross-cutting dynamic of their work with districts.

Several learning theorists have elaborated the notion of adaptation with regard to expertise (Hatano & Inagaki, 1986; Schwartz, Bransford, & Sears, 2005).<sup>2</sup> Like adaptive expertise, adaptive assistance involves activities—in this case, those previously elaborated—that participants engage in deeply but also break out of within certain parameters depending on dimensions of their situation such as district practitioners’ starting capacity. When assistance is adaptive, participants in those assistance relationships do not simply replicate behaviors of the past but continually assess their situations (especially the extent to which those situations are routine or nonroutine); take action; and revisit the fit between their goals, actions, and outcomes.

Virtually all IFL staff reinforced adaptive assistance as the overarching orientation to their district assistance relationships. In the words of one IFL staff person, “[our] best work is done when these lessons [about local constraints on their work] are heated and then our work is adapted to try to assist the district and support the people there to deal with or modify those constraints.” IFL fellows and other staff typically described the IFL as an “R&D” or research and development organization that aimed to support its staff in continually interrogating what they were learning, translating those lessons and other evidence into resources for districts, testing those resources in particular circumstances, and reinforcing or revising what they do.

Principals and central office administrators in our focal districts highlighted that IFL staff’s “responsiveness” and efforts “to tailor” their work to the specific interests and needs of districts helped them and others in their districts to sustain and deepen their engagement in the partnerships over time. For example, one central office administrator highlighted,

Well I would say that, as our capacity within the district has grown, it has gone more back and forth. Initially, it was primarily delivered from the IFL to us. But that’s part of why we went to them because they have that knowledge base. And as we started to grow— . . . I’ve seen evidence of IFL really adapting the work based on what they’re seeing on our end. And not just what they’re seeing, but some people—like I’m very vocal about what I want and don’t want or what I think works and doesn’t work. And I think if they weren’t responsive to that and respectful of that, the relationship wouldn’t continue. It doesn’t mean they do everything that I ask. They do what they do with reason then we have a healthy dialogue about why something should be one way or the other. And that’s not just me. They have it with our curriculum supervisors, with our coaches, with our principals.

---

<sup>2</sup>Other learning theorists call this kind of engagement double-loop learning (e.g., Argyris, 1976; Argyris & Schon, 1996) or trial-and-error learning (Levitt & March, 1988).



Another corroborated that their interactions with the IFL suggest their orientation toward continually revisiting and revising how they work with districts

if they didn't evolve, and I am talking about evolving, changing the way they think, I mean, that's what education is all about. It's the way—we need to change. I mean, if you have a child in front of you who's not getting it the way you're teaching it this way, then [you've] got to do it another way. And then you've got to try something else. And not that I'm saying that it wasn't working that way, but you can't do same old same old all the time, especially while people don't listen to same old, same old. . . . It's wonderful. [IFL staff] try—they do listen to feedback and they do try new things.

Our observations and document reviews suggested that tool development, in particular, reflected these characteristics of adaptive assistance. We found multiple instances of IFL staff continuing to refine tools as they applied them in new situations and collected new information about how the tool played out in practice in those situations. The IFL labeled various versions of its tools (e.g., Learning Walk Version 2.1) to mark this evolution and to institutionalize the regular practice of revising tools.

For example, by the end of our data collection, the Principals of Learning (POL) had evolved over a series of years into what some IFL staff had come to call a “suite” of tools. The suite included CD-ROMs containing an e-book with written explanations of the POLs, examples of student work, video clips of classroom practice that reflected the principles, and other materials designed to engage district practitioners in thinking about how the POLs related to their own practice. As one IFL fellow explained,

Over time, one of the things we wanted people to be able to see is the Principles of Learning living in a classroom; but then as we learned more about talk and content, we realized you can't really do a LearningWalk without knowing something about the content that you're looking at. So the LearningWalk had to change to reflect the fact that . . . you can't look at rigor absent the content of the subject matter. So the research on that informed how we had to think about the LearningWalk. And then actually . . . we also realized that we had to look at the research on trust . . . in order to have people feel that the LearningWalk is something that they were willing to be involved in, so we had to look at [Anthony] Bryk's work on trust. So over the course of time as we did the LearningWalk and realized there were pieces missing, we'd go back to research and think what parts of this are we not doing well enough that causes the kind of problems that the LearningWalk has? So are we being reflective on why this wasn't working, and what other pieces do we need to do.

Various comments from IFL fellows suggested that such adaptive assistance was typical of work across IFL staff. For example, one IFL staff person reported

how its Disciplinary Literacy Team (one of its development teams) members adapted general ideas about coaching to coaching in specific discipline areas (such as math, science, English language arts, and social studies):

So [in developing one set of tools] there was a time where we [the Disciplinary Literacy team] kind of put our heads together and we created a general session that they would engage in. . . . And then the team—every one of us would go off and put [the materials] into our [discipline-specific] work and we shared how we were going to do that. And then the team came back together again. . . . So we can really come out with—okay, what are we calling our coaching model because we can't each have our own definitions [of coaching for each discipline]; something has to tie us together as an organization.

IFL fellows also experienced a tension in engaging in adaptive assistance especially in the context of tool development: whether to keep a tool or another dimension of their work the same or revise it in light of new ideas. For example, some stability in their tools seemed essential especially in light of the significant length of time it could take a critical mass of district practitioners across a district to engage in any single iteration of a tool. However, engaging district practitioners in outdated versions of tools threatened the potential impact of their work. Accordingly, adaptive assistance in the context of tool development and ongoing tool use seemed to depend on IFL staff continually revising their tools but also maintaining some stability in their tools. This tension seemed most pronounced as IFL staff grappled with whether one district's experience with a tool should be generalized to others and used to remake a given tool or whether their experience was idiosyncratic or local.

## SUMMARY, CONCLUSIONS, AND IMPLICATIONS

In this article we draw on five years of data on the IFL's work with eight districts to build knowledge about learning-support intermediary organizations. The IFL offered an important case for this inquiry in part because past research associated the IFL with such outcomes as changing district central office administrators' and school principals' thinking and work practices in ways that seemed to enable the implementation of learning improvement efforts. However, this research did not elaborate what specifically about the IFL's work with districts might account for such improvements. We found that concepts from sociocultural learning theory helped capture main dimensions of the IFL–district assistance relationships that practitioner reports and observational data associated with such outcomes and that sociocultural learning theory suggests contribute to changes in work practice. These dimensions include brokering, modeling, the provision of certain social

opportunities, tool development, and engagement in joint work. We did not find evidence that the IFL staff created and valued identity structures as part of their assistance relationships. They did, however, deliberately structure activities with district practitioners in ways that promised to limit risks practitioners' may have associated with exploring new work practices. They also took a particular approach to tool development that involved integrating knowledge from both research and practice into tools over time. We argue that the IFL continually revisited how it engaged in these activities, especially tool development, in ways that seemed fundamental to their approach to their work. Accordingly, we characterize their assistance relationships as adaptive assistance. We focus here on what the IFL's work involved when it reflected these activities at a high level. But some dimensions of their work such as engagement in joint work seemed uneven over time.

This research has a number of implications for district leaders and other policy-makers. For one, the case of the IFL suggests that learning-support intermediaries may be important for helping achieve their learning-support goals and that they provide such support by engaging in district assistance relationships with particular features. By their own and other accounts, the IFL has not been a purveyor of a particular reform approach or focused only on schools like some external support organizations who operate like vendors or school coaches. Rather IFL staff members have tried to position themselves between central offices and schools as a responsive, engaged district partner with the capacity to bring a range of resources to bear on district-specific challenges and priorities for both school- and central office-level staff. Our conceptual framework, derived from studies of learning across multiple settings over time, suggests that such multilevel, situated supports are associated with deepening people's engagement in various challenging work practices. District leaders and other policymakers might examine whether they have access to external partners with the capacity to engage in the kinds of adaptive assistance relationships described here and whether such partnerships would enhance their own work.

District and other educational leaders might also consider that maintaining and growing intermediary organizations capable of the assistance relationships described here may demand that they make substantial investments in intermediary organizations over time. The IFL received seed funding from the University of Pittsburgh, ongoing support from foundations, and a demand for their services that generated a steady revenue source from districts. Other school reform support organizations may not have ready access to resources that would enable them to operate in ways that resemble the activities described here. Public and private funders in particular might examine how they can create funding opportunities for intermediary organizations to engage in multiyear, district-responsive relationships. Such grant making may require some funders to significantly reform their grant-making strategies—especially funders that traditionally have invested in the delivery of specific programs for discrete periods rather than in enabling dynamic,

locally responsive relationships between intermediary organizations and districts over time. Similarly, school districts might consider how they can procure funding for adaptive-assistance partnerships. Such investments may be a particularly hard sell for some school boards whose contracting and accountability mechanisms favor targeted technical assistance—in particular, predetermined areas, not adaptive assistance relationships.

For other external learning-support organizations, the IFL case provides one model for how they might go about the work of supporting districtwide learning improvement efforts. The IFL example is one of an intermediary organization—one that trains its efforts on shifting work practices within both central offices *and* schools and one that serves as a bridge between central offices and schools in the process. Our study reveals specific activities in which the IFL engaged in these in-between spaces. Although various organizations claim to develop so-called tools to support district improvements, the IFL's approach to integrating knowledge from research and practice into their tools, and to growing their tools over time, seemed to result in a particularly powerful set of resources for districts. Other organizations might consider whether their work might be improved by engagement in these types of activities. In the process, members of other learning-support organizations—and IFL staff, too—might consider whether the ways the IFL case deviated from theory might point to potential liabilities that practitioners should address in their future work. For example, would the work of the IFL and other organizations be enhanced if it included the development of identity structures for practitioners? Would these organizations provide stronger supports for districts if they bridged to research not only on how people learn but on leadership, implementation, change, and other areas that the IFL typically did not engage?

Our work also raises a number of questions for future research. For one, although we examined eight district partnerships over five years, our study is still an examination of one intermediary organization. Do the findings in the IFL case bear out in the context of other intermediary organizations? Additional confirming or disconfirming cases would greatly deepen the theoretical and empirical base about intermediary organizations and learning improvement.

In the process of pursuing other cases—including perhaps the IFL's relationships with other districts—researchers might focus their data collection specifically on elaborating the concept of adaptive assistance. As previously noted, such forms of participation come with fundamental tensions such as how to manage competing demands to direct and co-construct work. Also, as previously noted and elaborated elsewhere, IFL fellows grappled with whether particular forms of practitioners' engagement with a tool such as the LearningWalk protocol constituted appropriate or inappropriate use of that tool. Further research on "adaptive expertise" might seek to clarify how participants in these assistance relationships manage these tensions. Researchers might also consider how an observer of, or participant in,



these relationships might gauge when a given action reflects adaptive behavior (i.e., a variation within an appropriate range of behavior) versus maladaptive participation.

Future research might also aim to capture dynamics of intermediary work over time. As we previously noted, given the nascent stage of research in this area, we examined the extent to which IFL work frequently resembled features of assistance relationships described in our conceptual framework. However, such developmental and challenging work likely waxes and wanes over time as all participants in the relationships build their capacity for deeper engagement in particular endeavors. For example, in the early stages of the IFL's relationship with some districts, their activities departed distinctly from the reciprocal, nondirective nature of their relationships. Elaborating how such relationships change—or what sociocultural learning theorists might refer to as the trajectory of intermediary work—seems important to deepening knowledge about what these organizations are and what they do. What is the trajectory of this kind of work and how do such trajectories vary by district and intermediary starting capacity among other conditions? How can other cases and alternative research designs capture these intermediary dynamics?

The findings presented here, as well as our broader analysis of IFL–district partnerships (Honig & Ikemoto, 2006), suggest that particular conditions help and hinder intermediary work. For example, the IFL's fiscal resources seemed to provide foundational supports. As we suggest here but elaborate elsewhere, all IFL fellows interviewed for this study bring to their work significant knowledge resources, as well as experiences in district practice and in using research that enabled them to engage in challenging work with various practitioners over time. Conditions within districts such as accountability pressures may help or hinder intermediary–district partnerships, depending on how participants manage those pressures. Researchers would develop a fuller picture of what intermediary organizations do and how they do it if they elaborated the multiple contexts in which intermediary organizations operate and how features of those contexts mediate the work.

Ultimately, future research should probe whether and how specific activities of intermediary organizations contribute to learning improvements. We identified IFL activities associated with changes in district policy and how central office administrators and school principals thought about and engaged in their work. But to what extent do these changes in leadership practice translate into expanded learning opportunities and outcomes for students? Especially because intermediary work unfolds in between schools and central offices, and because it is highly contingent on the capacity and practices of others, linking intermediary activities with such outcomes will require significant conceptual development.

## REFERENCES

- Argyris, C. (1976). Single-loop and double-loop models in research on decision making. *Administrative Science Quarterly*, 21(3), 363–375.
- Argyris, C., & Schon, D. A. (1996). *Organizational learning II: Theory, method, and practice*. Reading, MA: Addison-Wesley.
- Berends, M., Bodilly, S. J., & Kirby, S. N. (2002). Looking back over a decade of whole-school reform: The experience of New American Schools. *Phi Delta Kappan*, 84(2), 168–175.
- Berends, M., Chun, J., Schuyler, G., Stockly, S., & Briggs, R. J. (2002a). *Challenges of conflicting school reforms*. Santa Monica, CA: RAND Corporation.
- Berends, M., Chun, J., Schuyler, G., Stockly, S., & Briggs, R. J. (2002b). Implementation of NAS designs in a high-poverty district. In M. Berends, S. J. Bodilly, & S. N. Kirby (Eds.), *Facing the challenges of whole-school reform: New American schools after a decade* (pp. 95–122). Santa Monica, CA: RAND Corporation.
- Bodilly, S. J., & Berends, M. (1999). Necessary district support for comprehensive school reform. In G. Orfield & E. H. DeBray (Eds.), *Hard work for good schools: Facts not fads in Title I reform* (pp. 111–119). Boston: The Civil Rights Project, Harvard University.
- Bodilly, S. J., Glennan, T. K., Jr., Kerr, K. A., & Galegher, J. R. (2004). Introduction: Framing the problem. In T. K. Glennan, Jr., S. J. Bodilly, J. R. Galegher, & K. A. Kerr (Eds.), *Expanding the reach of education reforms: Perspectives from leaders in the scale-up of educational interventions* (pp. 1–40). Santa Monica, CA: RAND Corporation.
- Brown, A., & Campione, J. (1994). Guided discovery in a community of learners. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 229–270). Cambridge, MA: MIT Press/Bradford.
- Brown, J., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32–42.
- Brown, J. S., & Duguid, P. (1991). Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation. *Organization Science*, 2(1), 40–57.
- Cobb, P., & Bowers, J. S. (1999). Cognitive and situated learning perspectives in theory and practice. *Educational Researcher*, 28(2), 4–15.
- Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, 23(2), 145–170.
- Coburn, C. E. (2005). Shaping teacher sensemaking: School leaders and the enactment of reading policy. *Educational Policy*, 19(3), 476–509.
- Cohen, D. K. (1982). Policy and organization: The impact of state and federal educational policy on school governance. *Harvard Educational Review*, 52(4), 474–499.
- Collins, A. M., Brown, J. S., & Holum, A. (2003). Cognitive apprenticeship: Making thinking visible. In L. B. Resnick & M. W. Hall (Eds.), *The principles of learning: Study tools for educators* (pp. 1–18). Pittsburgh, PA: University of Pittsburgh.
- Corcoran, T., Fuhman, S. H., & Belcher, C. L. (2001). The district role in instructional improvement. *Phi Delta Kappan*, 83(1), 78–84.
- Datnow, A., Hubbard, L., & Mehan, H. (2002). *Extending educational reform from one school to many*. New York: Routledge Falmer.
- Derry, S., Gance, S., & Schlager, M. (2000). Toward assessment of knowledge-building practices in technology-mediated work group interactions. In S. P. Lajoie (Ed.), *Computers as cognitive tools, Volume 2: No more walls*, (pp. 29–67). Mahwah, NJ: Erlbaum.
- Elmore, R. F., & Burney, D. (1998). *Continuous improvement in Community District #2, New York City*. Pittsburgh, PA: High Performance Learning Communities Project, Learning Research and Development Center, University of Pittsburgh.

- Engestrom, Y., & Miettinen, R. (1999). Introduction. In Y. Engestrom, R. Miettinen, & R.-L. Punamaki (Eds.), *Perspectives on activity theory* (pp. 1–38). Cambridge, UK: Cambridge University Press.
- Gallucci, C., Boatright, B., Lysne, D., & Swinnerton, J. (2006, April). *District reform as teaching and learning: How the system "learns" to improve instruction*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.
- Glennan, T. K., Jr., & Resnick, L. B. (2004). School districts as learning organizations: A strategy for scaling education reform. In T. K. Glennan, Jr., S. J. Bodilly, J. R. Galegher, & K. A. Kerr (Eds.), *Expanding the reach of education reforms: Perspectives from leaders in the scale-up of educational interventions* (pp. 517–563). Santa Monica, CA: RAND Corporation.
- Grossman, P. L., Smagorinsky, P., & Valencia, S. (1999). Appropriating tools for teaching English: A theoretical framework for research on learning to teach. *American Journal of Education*, 108(1), 1–29.
- Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, & K. Hakuta (Eds.), *Child development and education in Japan* (pp. 262–272). New York: Freeman.
- Hightower, A. M., Knapp, M. S., Marsh, J. A., & McLaughlin, M. W. (Eds.). (2002). *School districts and instructional renewal*. New York: Teachers College Press.
- Hill, H. C. (2001). Policy is not enough: Language and the interpretation of state standards. *American Educational Research Journal*, 38(2), 289–318.
- Holland, D., Lachicotte, W., Skinner, D., & Cain, C. (1998). *Identity and agency in cultural worlds*. Cambridge, MA: Harvard University Press.
- Honig, M. I. (2001). *Managing ambiguity: The implementation of complex education policy*. Unpublished doctoral dissertation, School of Education, Stanford University, Stanford, CA.
- Honig, M. I. (2004). The new middle management: The role of intermediary organizations in complex education policy implementation. *Educational Evaluation and Policy Analysis*, 26(1), 65–87.
- Honig, M. I. (in press). District central office administration as learning: How socio-cultural and organizational learning theories elaborate district-central-office participation in teaching and learning improvement efforts. *American Journal of Education*.
- Honig, M. I., & Ikemoto, G. S., (2006). *Making and re-making the link between research and practice: The case of the institute for learning*. Research report prepared for the MacArthur Network for Teaching and Learning, Stanford, CA.
- Hubbard, L., Mehan, H., & Stein, M. K. (2006). *Reform as learning*. New York: Routledge.
- Ikemoto, G. S. & Honig, M. I. (in press). Tools to deepen practitioners' engagement with research: The case of the Institute for Learning. In C. E. Coburn & M. K. Stein (Eds.), *Research and practice in education: Building alliances, bridging the divide*. New York: Rowman & Littlefield.
- Institute for Learning. (2003). *LearningWalk<sup>SM</sup> sourcebook*. Pittsburgh, PA: University of Pittsburgh.
- Kirby, S. N., Berends, M., Naftel, S., McKelvey, C., Bodilly, S. J., Chun, J., et al. (2002). Implementation of NAS designs during the scale-up phase. In M. Berends, S. J. Bodilly, & S. N. Kirby (Eds.), *Facing the challenges of whole-school reform* (pp. 71–94). Santa Monica, CA: RAND.
- Lave, J. (1996). Situating learning in communities of practice. In L. B. Resnick, J. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 63–82). Hyattsville, MD: American Psychological Association.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.
- Lee, C. D. (2001). Is October Brown Chinese? A cultural modeling activity system for underachieving students. *American Educational Research Journal*, 38(1), 97–141.
- Leithwood, K., & Louis, K. S. (Eds.). (1998). *Organizational learning in schools*. Lisse, The Netherlands: Swets & Zeitlinger.
- Levitt, B., & March, J. G. (1988). Organizational learning. *American Review of Sociology*, 14, 319–340.
- Marsh, J. A., Kerr, K. A., Ikemoto, G. S., & Darilek, H. (2006). Developing district intermediary partnerships to promote instructional improvement: Early experiences and lessons about the Institute for

- Learning. In K. Wong & S. A. Rutledge (Eds.), *System-wide efforts to improve student achievement* (pp. 241–270). Greenwich, CT: Information Age.
- Marsh, J. A., Kerr, K. A., Ikemoto, G. S., Darilek, H., Suttorp, M. J., Zimmer, R., et al. (2005). *The role of districts in fostering instructional improvement: Lessons from three urban districts partnered with the Institute for Learning*. Santa Monica, CA: RAND.
- McLaughlin, M. W. (2006). Implementation research in education: Lessons learned, lingering questions, and new opportunities. In M. I. Honig (Ed.), *New directions in education policy implementation: Confronting complexity* (pp. 209–228). Albany: The State University of New York Press.
- McLaughlin, M. W., & Talbert, J. (2001). *Professional communities and the work of high school teaching*. Chicago: The University of Chicago Press.
- Murphy, J., & Hallinger, P. (2001). Characteristics of instructionally effective school districts. *Journal of Educational Research*, 81(301), 175–181.
- Resnick, L. B., Hall, M. W., & Fellows of the Institute for Learning. (2001). *The Principles of Learning: Study tools for educators* [CD-ROM]. Pittsburgh, PA: University of Pittsburgh, Institute for Learning, Learning Research and Development Center.
- Rogoff, B. (1994). Developing understanding of the idea of communities of learners. *Mind, Culture, and Activity*, 1(4), 209–229.
- Rogoff, B., Baker-Sennett, J., Lacas, P., & Goldsmith, D. (1995). Development through participation in sociocultural activity. In J. Goodnow, P. Miller, & F. Kessel (Eds.), *Cultural practices as contexts for development* (pp. 45–65). San Francisco: Jossey-Bass.
- Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 1–51). Greenwich, CT: Information Age.
- Scribner, J. P., Cockrell, K. S., Cockrell, D. H., & Valentine, J. W. (1999). Creating professional communities in schools through organizational learning: An evaluation of a school improvement process. *Educational Administration Quarterly*, 35(1), 130–160.
- Scribner, J. P., Hager, D. R., & Warne, T. R. (2002). The paradox of professional community: Tales from two high schools. *Educational Administration Quarterly*, 38(1), 45–76.
- Smagorinsky, P., Cook, L. S., & Johnson, T. S. (2003). The twisting path of concept development in learning to teach. *Teachers College Record*, 105(8), 1399–1436.
- Snipes, J., Doolittle, F., & Herlihy, C. (2002). *Foundations for success: Case studies of how urban school systems improve student achievement*. Washington, DC: MDRC for the Council of the Great City Schools.
- Spillane, J. P. (2000). *District leaders' perceptions of teacher learning* (No. OP-05). Philadelphia: Consortium for Policy Research in Education.
- Spillane, J. P. (2002). Local theories of teacher change: The pedagogy of district policies and programs. *Teachers College Record*, 104(3), 377–420.
- Spillane, J. P., Diamond, J. B., Burch, P., Hallett, T., Jita, L., & Zoltners, J. (2002). Managing in the middle: School leaders and the enactment of accountability policy. *Educational Policy*, 16(5), 731–763.
- Spillane, J. P., & Thompson, C. L. (1997). Reconstructing conceptions of local capacity: The local education agency's capacity for ambitious instructional reform. *Educational Evaluation and Policy Analysis*, 19(2), 185–203.
- Stein, M. K., & Brown, C. A. (1996). Teacher learning in a social context: Integrating collaborative and institutional processes with the study of teacher change. In E. Fennema & B. S. Nelson (Eds.), *Mathematics teachers in transition* (pp. 155–191). Mahwah, NJ: Erlbaum.
- Swinnerton, J. A. (2006). *Learning to lead what you don't yet know: District leaders engaged in instructional reform*. Unpublished doctoral dissertation, University of Washington, Seattle.
- Tharp, R. G., & Gallimore, R. (1988). *Rousing minds to life: Teaching, learning, and schooling in social context*. New York: Cambridge University Press.



- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Weick, K. E. (1995). *Sensemaking in organizations*. Thousand Oaks, CA: Sage.
- Weick, K. E. (1998). Improvisation as a mindset for organizational analysis. *Organization Science*, 9(5), 543–555.
- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge, UK: Cambridge University Press.
- Wertsch, J. V. (1996). A sociocultural approach to socially shared cognition. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 85–100). Hyattsville, MD: American Psychological Association.
- Wertsch, J. V., del Rio, P., & Alvarez, A. (Eds.). (1995). *Socio-cultural studies of mind*. New York: Cambridge University Press.

# Authority, Status, and the Dynamics of Insider–Outsider Partnerships at the District Level

Cynthia E. Coburn, Soung Bae, and Erica O. Turner  
*University of California, Berkeley*

As school districts move toward systemic approaches to instructional reform, they are increasingly collaborating with outside organizations in this complex work. While emerging research touts the benefits of insider–outsider collaboration, we know little about the underlying processes by which partnerships are negotiated and maintained at the district level. Drawing on data from a longitudinal case study of a collaborative effort between an urban school district and a university-based research center, we investigate the role of authority and status in an insider–outsider partnership at the district level. We use conceptual tools from frame analysis and sociological theories of authority to describe the process by which authority and status relations develop. We then show that both authority and status shape how negotiation between insiders and outsiders unfolds. We argue that those with authority have a greater range of tools for negotiation and thus have greater influence. Status relations are important but are often mediated by authority relations. In addition, we argue that the organizational structure of the district shapes how the process unfolds in consequential ways. We conclude with implications for scholarship on and the practice of insider–outsider collaborations at the district level.

---

We thank Joan Talbert, Kristin Crosland, Angie Eilers, and Judith Toure for help with data collection; Mika Yamashita and Judith Toure for help with data analysis; and Grotius Hugo and Tara Amin for administrative assistance. We also thank the district personnel and outside support providers for letting us learn from their process, for their honesty about the challenges they faced, and their persistence in the face of difficult circumstances. Finally, we thank fellow researchers on the Meta Study for ongoing conversations about the dynamics of co-construction as well as Joan Talbert, Warren Simmons, and the research group on policy implementation at UC Berkeley for their thoughtful comments on earlier drafts of this article.

Correspondence should be sent to Cynthia E. Coburn, 3643 Tolman Hall, Berkeley, CA 94720.  
E-mail: cecoburn@berkeley.edu

Now more than ever before, school districts are attempting ambitious reform initiatives intended to improve instruction in schools throughout the district. As school districts move toward systemic approaches to instructional reform—as they attempt to foster instructional improvement at scale—they are increasingly reaching out to a range of external service providers to support them in this ambitious task (Burch, 2002; Gamoran et al., 2003; Glennan & Resnick, 2004; Honig, 2004; Marsh et al., 2005). Indeed, existing evidence suggests that collaboration with outside service providers can have positive outcomes for districts including increased capacity (Gamoran et al., 2003; Marsh et al., 2005) and greater access to research-based resources (Corcoran & Rouk, 1985; Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006; Spillane & Thompson, 1997).

In spite of this optimism about partnerships, research on insider–outsider collaboration in education suggests that establishing productive relationship can be fraught with difficulty. Different parties can come to the table with different priorities and agendas (Firestone & Fisler, 2002; Goodlad & Sirotnik, 1988; Heckman, 1988; Kornfeld & Leyden, 2001; Vozzo & Bober, 2001). Differences in status between researchers and practitioners can lead to tensions and conflict (Bickel & Hatrup, 1995; Freedman & Salmon, 2001; Goodlad & Sirotnik, 1988; Osajima, 1989). Unclear or unfamiliar roles and relationships on both sides can create uncertainty and misunderstanding (Freedman & Salmon, 2001; Goldring & Sims, 2005; Handler & Ravid, 2001; Hasslen, Bacharach, Rotto, & Fribley, 2001).

To date, most research on the dynamics of partnership has been done at the school level. Thus, we know little about the underlying processes by which partnerships are negotiated and maintained at the district central office level. Furthermore, although existing research has highlighted the role of unequal status in shaping collaborative relationships, it has paid little attention to the role of authority relations. Yet those working in a district central office are embedded in a web of complicated authority relations that characterize complex organizations. And those outside of the district have an uncertain position with regard to district authority relations.

Here, we draw on data from a longitudinal study of a collaborative effort between a midsize urban school district and a university-based research center to investigate the role of authority and status in insider–outsider partnerships at the district level. We draw on frame analysis and sociological theories of authority to investigate the dynamics of negotiation between outsiders and insiders as they set strategic priorities for their work with one another. In so doing, we uncover the process by which authority relations and status attributions develop in a partnership, arguing that they are situational and evolve over time. We further argue that both authority relations and status, once established, are crucial because they shape the microprocesses of negotiation between insiders and outsiders. Authority relations are especially important because those with formal or informal authority have a greater range of tools that they bring to the negotiation and thus have greater influence. We further show that the organizational structure of the district shapes

how the process unfolds in consequential ways. We close with implications of this research for scholarship on and the practice of insider–outsider collaborations at the district level.

## LITERATURE REVIEW

Existing research on insider–outsider partnerships is replete with the challenges involved in creating productive working relationships. Tensions around whose knowledge is valued can emerge as outsiders' knowledge is often accorded greater status in the culture at large than practitioner knowledge, especially if the outsiders are researchers or academics (Bickel & Hatrup, 1995; Gifford, 1986; Goodlad & Sirotnik, 1988; Osajima, 1989; Sinclair & Harrison, 1988). University researchers and school people also come from distinct cultures with different work practices, incentives, and senses of urgency about their work (Bickel & Hatrup, 1995; Brookhart & Loadman, 1992; Gifford, 1986; Goodlad & Sirotnik, 1988; Keating & Clark, 1988; Schlechty & Whitford, 1988). In addition, a history of poor relationships between academics and schools can make trust difficult to establish in new school–university partnerships (Gates-Duffield & Stark, 2001; Gifford, 1986; Hasslen et al., 2001; Lieberman, 1988; Rosen, 2008; D. D. Williams, 1988).

Existing research on insider–outsider partnerships has focused on collaborations at the school level (e.g., Boostrom, Jackson, & Hansen, 1993; Erickson & Christman, 1996; Firestone & Fisler, 2002; Lieberman, 1988; Osajima, 1989; Ravid & Handler, 2001; Sinclair & Harrison, 1988) or outside of the school or district context in university-based curriculum development projects or task forces (e.g., Bickel & Hatrup, 1995; Heckman, 1988; Keating & Clark, 1988; Lieberman, 1988; D. D. Williams, 1988). This work has yet to investigate the dynamics of insider–outsider collaboration at the district central office level. The district central office is much more complex organizationally than a school. Decision making at the district level is often stretched across multiple levels and multiple divisions, involving those with different levels of authority (Coburn, Honig, & Stein, in press; Spillane, 1998). Studying insider–outsider partnerships at the district level thus creates the opportunity to more fully understand the role of organizational structure in influencing how collaborative efforts unfold.

Furthermore, although existing research on insider–outsider relationships has focused a great deal of attention on the role of status, it has paid little attention to the role of formal and informal authority. Authority relations between those outside the system and those inside the system are at best uncertain. Technically, those in the district have formal authority over any given initiative under their jurisdiction. Those outside schools do not. Yet we know little about how authority relations are negotiated in the face of uncertainty and how they influence the



process of insider–outsider negotiation. Nor do we understand the relationship between authority and status in these partnerships.

Finally, most of the published writing on insider–outsider collaboration is not research. Rather, the literature is filled with reflective pieces written by the researchers or the practitioners involved in the collaboration (Bickel & Hattrup, 1995; Firestone & Fisler, 2002; and Rosen, 2008, are exceptions here). Although these pieces provide much insight into the key factors influencing the dynamics of partnership, they do not investigate the role of these factors systematically.

To understand the role of authority and status in insider–outsider partnerships at the district level, we focus attention on the dynamics of negotiation between outsiders and insiders as they set strategic priorities for their work with one another. When insiders and outsiders come together to collaborate on a new initiative, they often come with conflicting ideas about the direction they should take their shared work (Gifford, 1986; Heckman, 1988; Keating & Clark, 1988). Different visions of the problems that need to be addressed or appropriate solutions to pursue must be resolved in order to move forward. We draw on frame analysis and theories of authority from organizational sociology to understand the dynamics of this negotiation.

Frame analysis represents a set of conceptual tools for investigating the way ideas are produced and invoked to mobilize people to action. It helps us understand the process by which people come to understand the nature of the problem and potential solutions through social interaction and negotiation. Thus, in the case of insider–outsider collaboration, frame analysis helps us understand how directions for joint work get negotiated as individuals from districts work with those from the outside over time. Frame analysts identify two kinds of problem frames that individuals and groups invoke in their on-going interaction: diagnostic and prognostic (Benford & Snow, 2000; Snow & Benford, 1992). Diagnostic framing involves defining problems and attributing blame. How a problem is framed is important because it focuses attention on some aspect of the problem and not others, identifies some individuals or groups as responsible for the problem, and thus identifies those responsible for change (Cress & Snow, 2000; Stone, 1988). Prognostic framing involves articulating a proposed solution to the problem. In so doing, a prognostic frame puts forth particular goals and suggests tactics for achieving those goals (Benford & Snow, 2000; Cress & Snow, 2000; Snow & Benford, 1992). Diagnostic and prognostic framing are often closely intertwined, as prognostic framing often rests implicitly on problem definition and attribution that is part of diagnostic framing.

The act of framing is an interactive one constituted by two related processes: frame alignment and resonance. *Frame alignment* refers to the actions taken by those who produce and invoke frames to connect these frames with the interests, values, and beliefs of those they seek to mobilize (Snow, Rochford, Worden, & Benford, 1986; R. H. Williams & Kubal, 1999). Individuals and groups attempt to

construct ways of framing the problem that provide “conceptual hooks” (Zucker, 1991) that allow targets of mobilization to link the frame with other things they know, experience, and/or believe (Benford & Snow, 2000; Snow et al., 1986). But frame alignment activities are always dependent on how the individuals and groups respond, or what frame analysts call *resonance* (Snow et al., 1986; R. H. Williams & Kubal, 1999). Resonance is the “mobilizing potency” of a particular frame: the degree to which a frame is able to create a connection—a “deep responsive chord” (Binder, 2002, p. 220)—with individuals and motivate them to act.

Framing is often a contested process. Prognostic and diagnostic framing may be challenged as others offer counterframes that put forth alternative portrayals of the situation, often with contrasting implications for roles, responsibility, and resources (Benford & Snow, 2000; Fligstein, 2001; Stone, 1988). These frame disputes, as Benford and Snow called them, may stretch over time as frames are reconstituted and reframed in negotiation and interaction (Davies, 1999; Gamson, 1992). Furthermore, this negotiation among and between frames is likely to be shaped by relations of authority (Coburn, 2006; Fligstein, 2001; Isabella, 1990).

However, although some frame analysts acknowledge the role of authority in the problem framing process, few investigate it explicitly. Thus, the relationship remains undertheorized. For this, we turn to Dornbusch and Scott’s work on authority from organizational sociology. Authority can be defined as legitimized power relations (Dornbusch & Scott, 1975; Pace & Hemmings, 2007). In any social relationship, whether it is in formal organizations or informal group settings, relations of power and control come to be legitimized by rules and social norms (Dornbusch & Scott, 1975). Authority can be *authorized* as when those higher up in the organizational structure grant power to certain individuals. In this case, authority is power that is sanctioned by norms from above. But authority can also be *endorsed*, as when power relations are defined and enforced by those who are subject to the exercise of that power (Dornbusch & Scott, 1975; Scott & Davis, 2007). Authority relations in a given setting are likely to be most stable when they are simultaneously authorized from above and endorsed from below. However, in the absence of agreed-upon norms legitimizing power relations (either authorized or endorsed), authority relations fail to materialize. In the absence of clear authority relations, joint work is characterized by conflict, power struggles, and an inability to move forward (Dornbusch & Scott, 1975).

Authority can be formal or informal. Formal authority is power that is “coded into structural design” (McAdam & Scott, 2005, p. 10). That is, it is the authority that comes with a particular role or position in an organization and can be exercised by any person holding that position (Scott & Davis, 2007). Thus, in insider–outsider partnerships at the district level, district leaders have formal authority over people involved in any initiatives that emerge from the collaboration, although their degree of formal authority depends upon where they are in the district hierarchy. Outsiders do not have formal authority over the individuals they work

with in the district. Informal authority, by contrast, is authority that is acquired by an individual that is related to some special characteristics, such as specialized expertise or their position in a social network (Scott & Davis, 2007). Both insiders and outsiders can be accorded informal authority if they are either authorized to lead by those who have formal authority or endorsed by those who do not.

Status is also negotiated in social interaction. Individuals grant status to others in a social setting when they perceive that they have specialized expertise or skill that they can bring to bear in ways that benefit the joint work (Dornbush & Scott, 1975; Scott & Davis, 2007). Balkwell (1994) called status “unobserved performance expectations” (p. 124) that often result in power and prestige in group interaction. At times, individuals grant status in groups based on characteristics that are valued in the larger society, such as class background, race, or gender (Cohen, 1994). Sociologists call this phenomenon ascribed status. Prior research on insider–outsider partnerships suggests that individuals also grant others status based on occupational prestige or academic background—a form of what is known as achieved status—leading to greater influence for university researchers in the dynamics of partnership (Bickel & Hatrup, 1995; Goodlad & Sirotnik, 1988; Osajima, 1989; Sinclair & Harrison, 1988). It is important to note that those who are perceived to have greater status in a group—either ascribed or achieved—may be granted informal authority, either through authorization or endorsement (Cohen, 1994; Dornbush & Scott, 1975). But, status alone does not lead to greater authority in the absence of normative agreement from above or below that the person with status warrants greater authority in the collaboration.

Preliminary work on the role of authority in frame dynamics suggests that those with formal authority have greater influence in frame debates than those without formal authority. Individuals in positions of authority have greater access to others and can use this access to make their case. They also can control the focus of discussion or the agenda, and they often have the ability to control who participates in the decision process. Those with formal authority are able to use these features of their position to leverage their ideas, thus supporting their ability to persuade others of the wisdom of their view of the problem and prescription for solutions (Coburn, 2006; Coburn, Toure, & Yamashita, in press). However, even with those advantages, individuals with formal authority are not always able to persuade others of their position (Coburn, 2006). In this case, they may resort to more direct uses of authority, such as compelling others to act (Coburn, Toure, & Yamashita, in press). Finally, framing activities—especially frame disputes—can be occasions where authority relations are renegotiated and reshaped as well (Coburn, 2006).

Yet there is still much to learn about the role of authority, its relation to status, and the influence of both on the framing process. Here, we add to the research on insider–outsider relationships and research on framing in three ways. First, we uncover the dynamics by which authority relations are developed, paying



careful attention to how the philosophy of partnership influenced how participants constructed roles in relation to one another. Second, we illustrate how authority relations and status influence the tactics that individuals use while framing arguments and the degree to which these tactics are successful. Finally, we bring organizational structure into the equation, showing how the structure of the district creates the conditions within which authority relations develop, shift, and are renegotiated over time.

## METHODS

To understand the role of authority in insider–outsider negotiation, we draw on data from a longitudinal case study of one midsize urban school district involved in a partnership with an outside support provider. At the time of the study, the district served approximately 50,000 students, the majority of whom were low-income students of color and one fourth of whom were classified as English Language Learners. The partnership—which we call Partnership for District Reform (PDR)<sup>1</sup>—brought together members of a university research center and the school district to join research knowledge with clinical expertise in support of continuous instructional improvement at scale.

### *PDR*

According to the tenets of PDR, the collaborative work in the initiative was guided by the principle of co-construction, which called for district and external partners to collaboratively identify problems and develop and implement solutions that would be informed by research but adapted to local conditions and capacities. This approach emphasized the importance of both research knowledge and clinical knowledge for solving the problems the district faced. It was to be a partnership where diverse forms of knowledge were valued, stakes were shared, and differences of opinion were adjudicated with reference to evidence. Thus, in many ways, this initiative sought to address the status problems identified by prior research on insider–outsider collaborations by intentionally and publicly granting equal status for diverse forms of knowledge.

The outside research center coordinated a large number of external partners who came to the district to participate in this endeavor, including researchers from the research center, professors from several local universities, and experienced practitioners who were working as national consultants. In the second year of PDR, a second organization—a national organization devoted to district systemic

---

<sup>1</sup>Partnership for District Reform is a pseudonym.



change—was brought on board to provide additional capacity to support the initiative. On the district side, PDR involved district personnel at multiple levels of the district, including the superintendent, assistant superintendents, directors of key divisions, and professional development providers in the division of curriculum and instruction. Thus, PDR took care to actively involve key individuals at the uppermost levels of the district, something that is frequently called for by the literature on insider–outsider relationships (Gifford, 1986; Goodlad & Sirotnik, 1988; Sinclair & Harrison, 1988).

During the initial years of the partnership, insiders and outsiders worked together on a number of interrelated initiatives, including redesigning the district's system of professional development to provide more coherent and sustained approaches to fostering teacher learning in reading and mathematics, the creation of frameworks in mathematics and reading to guide district policy making and professional development, and the preparation of a plan for coordinated leadership development, to name a few.

### *Research Design and Data Collection*

The database for this study emerges from two interlocking research projects that studied PDR from its inception in fall 2002 until spring 2005. The first author of this study led a team of researchers who studied PDR as part of a broader research project that sought to understand the relationship between research and practice in a range of school improvement efforts. We were not participants in the insider–outsider partnership itself nor were we evaluators of PDR. Rather, we were funded to investigate the dynamics of this partnership as they unfolded over time. We collaborated with a research team led by Joan Talbert at Stanford University that was funded to document the progress of the PDR project and provide formative and summative feedback to the district, the university collaborators, and the foundation that funded the initiative.<sup>2</sup> The two research projects collaborated on research design, protocol development, and data collection to ensure that research activities met the goals of both projects while minimizing burden to the site.

The joint research effort relied on in-depth interviewing (Spradley, 1979), sustained observation (Barley, 1990), and document analysis. Over the course of three years, researchers from the two teams conducted 71 interviews with 38 members of the central office and 3 union officials. We also conducted 36 interviews with 19 external partners who were working on the project in some capacity during the time of the study. As a supplement to this data, our research

---

<sup>2</sup>The foundation that funded the initiative and these two research projects prefers to remain anonymous to protect the confidentiality of the school district involved in the study. We are grateful for their support.

team conducted an additional 9 interviews with 8 members of the central office during follow-up data collection in the 2006–07 school year. All interviews were audiotaped and transcribed. The two teams supplemented the interviews with observations of 33 planning meetings. These meetings were at multiple levels of the central office from executive leadership meetings to planning meetings at the department level to project design meetings between district staff and external consultants. The observations were recorded with detailed fieldnotes, but on some occasions key meetings were audiotaped and transcribed. In addition, members of the two research teams observed 36 days of professional development for teachers and school leaders. This provided insight into both the fruits of the collaboration and the ways in which experience doing the professional development fed back into ongoing deliberation. Finally, numerous documents related to the partnership were collected and analyzed. These documents include minutes and agendas of meetings, draft and final copies of policy and planning documents as well as written feedback provided on draft documents, and reports to the funder from the district and the external partners.<sup>3</sup>

### *Data Analysis*

All data were entered into NUD\*IST, a software program for qualitative data analysis. We began our analysis by identifying seven instances of collaboration within the overall initiative. Each instance had different foci and mission. Each also involved a different, although at times overlapping, configuration of actors. (See Table 1 for a description of each instance of collaboration.) We reviewed our complete corpus of data to identify all data that were relevant to each instance, and we created a longitudinal record of the interaction between insiders and outsiders for it. Next we developed a coding scheme rooted in prior theory and then elaborated and extended in dialogue with the data using the constant comparative method (Strauss & Corbin, 1990). We were particularly interested in coding cognitive aspects of the collaborative process (including conceptions of high quality professional development, conceptions of high-quality instruction, conceptions of leadership, conceptions of “research based,” and conceptions of partnership), organizational aspects of the collaborative process (including authority, status, resources, linkages, trust, and staff turnover), and political aspects of the process (including politics of race, politics of language, and politics of instruction). The three authors of this article coded all data for one entire instance together (15% of the overall data) to develop interrater reliability. The rest of the instances were

---

<sup>3</sup>In addition to the district-level research activities described here, the two research teams also conducted longitudinal analysis in 10 case study schools in the district. We do not draw on these data in this article.

TABLE 1  
Description of the Instances of Collaboration

	Description	Key Actors
Overall initiative	Discussions and decision-making at the executive level about the direction of the PDR. Those involved in this work negotiated how to focus their efforts at the overall initiative level and how various PDR activities would be developed and implemented.	<ul style="list-style-type: none"> <li>■ Superintendent</li> <li>■ Assistant superintendents</li> <li>■ Executive council</li> <li>■ Departmental directors</li> <li>■ Three members of external research center</li> </ul>
Leadership development	Work focused on the development of a coherent leadership program for school leaders in the district. This included planning summer institutes for leadership and investigating different approaches to leadership development used in other districts to guide their own work.	<ul style="list-style-type: none"> <li>■ Assistant superintendents</li> <li>■ School principals</li> <li>■ Leadership council</li> <li>■ Two member of external research center</li> <li>● An external consultant</li> </ul>
Research and assessment	Work focused on using student assessment data to improve academic achievement in the district. This included a project to make individual student data and classroom data available to teachers through the internet and a project to identify district teachers who were consistently raising student test scores and learning about what practices these teachers used to ensure their success.	<ul style="list-style-type: none"> <li>■ Director of district's research office</li> <li>■ Staff of district research office</li> <li>■ Two members of external research center</li> </ul>
Literacy framework	Work focused on the development of the district's central policy document on literacy. The framework was intended to guide the district's efforts towards developing teachers' understandings of teaching literacy.	<ul style="list-style-type: none"> <li>■ Assistant director of professional development</li> <li>■ District staff developers</li> <li>■ Expert teachers</li> <li>■ Two members of external research center</li> <li>■ Four external consultants</li> <li>■ National panel of reading experts</li> </ul>

(Continued on next page)

TABLE 1  
Description of the Instances of Collaboration (*Continued*)

	Description	Key Actors
Literacy institute	Work focused on the design and development of a weeklong professional development experience for teachers on literacy. The district offered the summer institute over successive years to ensure the eventual participation of all the district teachers. The goals of the summer institute were to highlight current research and evidence-based practices.	<ul style="list-style-type: none"><li>■ District executive administration</li><li>■ Director of curriculum unit</li><li>■ Assistant director of professional development</li><li>● District staff developers</li><li>■ Expert teachers</li><li>■ Three members of external research center</li><li>■ Five external consultants</li></ul>
Math framework	Work focused on the development of the district's central policy on mathematics education. The framework was intended to bring together national and state standards, district curriculum, and grade level expectations in order to guide the district and teachers in improving math instruction.	<ul style="list-style-type: none"><li>■ Assistant director of professional development</li><li>■ Director of math</li><li>■ District staff developers</li><li>■ Expert teachers</li><li>■ One member of external research center</li><li>■ National panel of math experts</li><li>■ Two external consultants</li></ul>
Math institute	Work focused on the design and development of a weeklong professional development experience for teachers on math. The district intended to offer the summer institute over successive years to ensure the eventual participation of all the district teachers. The goals of the summer institute were to highlight current research and evidence-based practices.	<ul style="list-style-type: none"><li>■ District executive administration</li><li>■ Director of curriculum unit</li><li>■ Assistant director of professional development</li><li>■ Director of math</li><li>● District staff developers</li><li>■ Expert teachers</li><li>■ One member of external research center</li><li>■ Four external consultants</li></ul>

*Note.* PDR = Partnership for District Reform.



coded by a single coder, although we engaged in periodic spot checks of coding to ensure that consistency of coding was maintained throughout the coding process.

After all the data were coded, we analyzed how each dimension shaped the dynamics of framing during deliberation and debate. This initial analysis suggested that authority relations and status were particularly important. Therefore, we re-analyzed the data to get a more precise understanding of the roles these factors played. Because each instance of collaboration involved multiple actors in different stages of the process and because status and authority relations shifted over time, it would be imprecise to analyze authority relations and status for an instance as a whole. Instead, we opted to identify key decision points in each instance and analyze the particular configuration of status and authority of the individuals involved at each decision point. This strategy allowed us to do a more fine-grained analysis of the role these two factors played in the dynamics of negotiation.

First, we identified those with formal and informal authority in the configuration of actors in each decision point. To establish formal authority, we relied on the organization chart along with interview data that provided information about formal roles and responsibilities. To establish informal authority, we relied on interview data to assess the degree to which there was normative agreement that particular individuals should play a particular role. In the absence of agreement of all involved, we did not consider an individual to have informal authority. We also paid attention to the presence of power struggles and breaches, where individuals acted in ways that violated others' sense of appropriate action. We took these things as indications that authority relations had failed to materialize or existing authority relations were being contested.

To analyze status relations, we identified instances where individuals accorded achieved status to others in the collaboration—that is, when they viewed a given individual as having resources that were particularly valuable to the joint work in which they were engaged. We paid particular attention to the criteria by which individuals accorded such status. In our data, individuals accorded status to those that they perceived to have specialized skill or expertise rooted in professional or personal experience or garnered through academic or other training. After analyzing the criteria by which others were seen to have status in the collaboration, we then analyzed who in the partnership was seen to have status, along what dimensions, in what context, and according to whom.

We then analyzed how authority relations and status influenced frame dynamics at each decision point. We paid particular attention to the kinds of frame tactics that individuals with different forms of status and authority used in deliberations and the success of these tactics. We used a series of matrices (Miles & Huberman, 1994) to analyze patterns across decision points within a given instance and then compared patterns across instances.

We were also interested in how participants' beliefs about the given topic that was the focus of the collaboration shaped frame dynamics. For each instance of collaboration, we identified the key foci of discussion. We then created typologies of beliefs of the participants for each main foci. To create the typologies, we drew deductively from existing research and inductively from our data to capture the range of beliefs. Ultimately, we created the following typologies: conceptions of high-quality professional development, high-quality instruction in mathematics, high-quality instruction in literacy, what constitutes good research, and appropriate approaches to leadership development. We then drew on interview data to analyze where all individuals involved in a given instance fit on the relevant typologies. This analysis allowed us to ascertain the degree to which frame dynamics were playing out in contexts of shared or diverse beliefs. It also lent insight into when and under what conditions solution frames were persuasive to others. Again, we used a series of matrices to analyze and confirm patterns across instances.

## AUTHORITY RELATIONS AND STATUS DYNAMICS

As suggested by Firestone and Fisler (2002), organizations involved in collaborative relationships are not unitary. Rather, they are collections of individuals and subgroups, each with their own characteristics, resources, and expertise. Indeed, this was the case with the PDR project. Each instance of collaboration involved multiple individuals—both insiders and outsiders—in different aspects of the discussion at different times. Deliberations about the work moved up and down the system as the broad parameters for the direction of the work were negotiated between insiders and outsiders at the executive level and were subsequently elaborated, adapted, and at times transformed during insider–outsider collaboration at the lower levels of the system. Authority and status relations were central to the way that these negotiations played out at multiple levels of the system. But these relations varied according to the particular configuration of individuals involved in a particular aspect of the process.

In this section, we analyze the nature of authority relations and status in different aspects of the collaborative work. We begin with authority relations, arguing that they are contextual, evolve over time, and develop through a variety of routes. We then discuss the dynamics of status. We argue that status relations are much more complicated than prior scholarship would suggest. There are multiple criteria for granting status that are at work simultaneously and attributions of status to an individual are often quite specific; thus, status might be accorded to an individual along one dimension but not others. Ultimately, we show that in spite of this complexity, insiders tended to have greater authority and outsiders tended to have greater status in negotiations.

### *The Dynamics of Authority Relations*

The language of “collaboration” or “partnership” often obscures issues of authority. Yet scholars who study group processes suggest that authority relations are likely to emerge as individuals work with one another, even in informal settings and temporary collaborations (Dornbusch & Scott, 1975; Wheelan, 1994). Indeed, authority relations emerged in all but three instances of collaboration. In fact, clear authority relations actually enabled productive working relationships. In the absence of established authority relations, the interaction was characterized by conflict, mistrust, and the inability to get work done.

There were three principal ways that authority relations were established in insider–outsider groups. First, in some instances, authority relations were established *contractually* as part of the terms of employment for the outside partners. In spite of the fact that PDR advocated a partnership characterized by co-construction where all partners jointly set the terms of their work together, when school district leaders took responsibility for identifying and hiring outsiders to work on the project, they often hired them under terms that established a much more traditional consultant relationship. In the traditional consultant relationship, authority is held by insiders who establish priorities for joint work and can take or leave any advice or ideas that the outsider offers.

For example, one of the central goals of the second year of the project was to develop an overarching framework for mathematics to guide policymaking and professional development around mathematics instruction for the district. The district mathematics staff took the lead and hired external consultants to do the extensive work of crafting the framework. Under the terms of the consulting contract, district mathematics leaders set the parameters, but the consultants produced drafts of the frameworks for the district leaders to review, and revised them in light of district feedback. In this instance and others like it, roles and authority relations rooted in a traditional consultant model were agreed upon in advance and were clear to all involved. And, when the district leaders were unhappy with the performance of the consultants or the direction they were advocating as they were in two cases, they fired them or did not invite them back to work with the district.

The second way that authority relations were established was when someone with formal authority *authorized* an individual—insider or outsider—to take the lead on a particular initiative, thus granting him or her informal authority over others involved in the work. In this approach, those with formal authority let others know that it was his or her expectation that a particular person would play a leadership role in the work. For example, in the first year of the initiative, the superintendent authorized a senior member of the research center to lead in the design and development of the summer institutes, even though this individual, as an outsider, had no formal authority over anyone in the district. In consultation



with the director of the curriculum office, this senior member of the research center hired external personnel to work in a collaborative design process with district staff and expert teachers. He set expectations for how the process should unfold and articulated desired outcomes, reviewed and signed off on plans, communicated with the senior administration about the progress of the planning, and also mediated disputes that arose between insiders and outsiders in what was, at times, a stressful and challenging process. As suggested by Dornbusch and Scott (1975), granting authority via authorization was most likely to create stable authority relations when authority was also endorsed from below. In this example, not only was the senior leader of the research center authorized by the superintendent, but those involved in the design of the summer institutes uniformly saw him as the legitimate leader of the work. For example, those in the district—including those quite high up in the department of curriculum and instruction—consistently chose to run key decisions by this senior member of the research team to make sure he approved.

The third and most common way that authority relations were established in PDR occurred when normative agreement on authority *emerged* as a result of interaction among insiders and outsiders. Rather than being established in advance by contract or authorization, authority relations were negotiated among the individuals involved in the collaboration through the process of doing the work. Over time, roles were gradually defined in relation to one another, and some individuals came to be seen by others as having greater authority over key decisions.

In PDR, emergent processes led to quite varied authority relations between insiders and outsiders. For example, in the collaboration around leadership development, several members of the research center worked with an assistant superintendent to craft plans for districtwide professional development for school leaders. In spite of initial understandings that they would co-construct the work with one another, members of the research center reported in interviews that they felt the assistant superintendent wanted them to act as staff to the initiative, rather than share the leadership with her as they expected. According to an external consultant, the assistant superintendent “insisted on controlling [the leadership work]. . . . They wanted [an external consultant] to write stuff and give it to them.” Yet, although the outsiders were not happy with this arrangement as it evolved, they accepted it and the collaboration moved forward, governed by this set of authority relations.<sup>4</sup>

---

<sup>4</sup>Dornbusch and Scott (1975) made an important distinction between the perceived validity of authority relations and perceived propriety of authority relations. It is possible to believe that the authority relations are appropriate (perceived validity), without personally liking the authority relations as they have developed (perceived propriety). In this instance, there appeared to be normative agreement about the appropriateness of this set of authority relations (validity), even though outsiders involved in the relationship did not much like them (propriety).



In contrast, a different set of authority relations emerged in the collaboration between outside partners and the district research office as they worked on a project to identify teachers in the district who had better than expected test results with poor students and students of color and use these teachers as demonstration teachers from whom others in the district could learn. In this instance of collaboration, the director of research initially spoke about needing a particular external partner to sign off on his plans and later said that he depended too much on this person in the early stages of the work. He expressed, "There was never a sense that [the external partner] was dictating anything in what we did. If anything, I might have depended on him too much because he was so knowledgeable to kind of come up with the next steps." Eventually, the relationship evolved away from this dependency relationship such that they discussed all major decisions and did not move forward until both the external partner and the director of the district research department felt comfortable with the direction they were going. Thus, their partnership evolved into an arrangement of shared authority.

In PDR, it was common for authority relations to be established primarily through emergence, especially in the first year. This may be because the philosophy of co-construction, which set the parameters within which the collaboration unfolded, was silent on authority relations. Under the project's version of co-construction, individuals with different knowledge were to bring their knowledge to bear on pressing problems of practice. When there was a difference of opinion, the differences were to be resolved with reference to evidence. But, the theory did not specify norms of appropriate authority relations in this process. Furthermore, this is an unfamiliar form of partnership for school districts (Bryk, Rollow, & Pinnell, 1996). Indeed, as we see, many of the people involved in the initiative, including some of the outside partners affiliated with the research center, were uncertain about appropriate roles and relationships under the theory of co-construction.

At times, this emergent process was quite bumpy. In three instances, there were moments where the collaboration was marked by struggles for control or by what ethnographers call breaches (Feldman, 1995), whereby one partner acted in ways that violated others' sense of appropriateness, leading to conflict. For example, during the first year of the initiative, the collaboration to design summer professional development in literacy was particularly challenging. The overall process was led by a senior member of the research center who had been authorized by the superintendent. This individual hired external consultants—some of whom were academics at a neighboring university and others of whom were former or current practitioners—to work with the district professional development staff and experienced teachers to collaboratively design a series of weeklong summer institutes for district teachers. They were told that they were to co-construct the institutes.

However, members of the district and external consultants came to the table with very different understandings of high-quality professional development and had very different ideas about the approach they should take in the joint work. For example, one district staff development provider described the difference between outside partners from the university and the views of the district professional development providers in the following way:

The University doesn't understand our audience and . . . [doesn't] realize that this has to be concrete, real, practical, take it tomorrow and use it, really use. And the University is really good at making you think about what you're doing and reflect, but we wanted more real experiences for the institute that teachers could model their instruction after, and not so much heady thinking time, but more: This is a technique. This is a method. This is an approach. This is the way. This is a model of how you would do this strategy.

In contrast, outside partners argued for a quite different model of professional development:

[It] needs to be a thoughtful and careful combination of talking about hard issues in reading instruction and something useful. By useful, I mean it could be ways to look at your classroom data or ways of looking at texts to determine the appropriate level for text selection. But in doing the useful things, [you] need to tie it back to why these things are important and underlying conceptual issues so it's not just: This is what you need to do.

Discussions about the appropriate approach to professional development stretched across multiple meetings with little movement on either side. Tensions rose and relations of trust began to fray. In part, insiders and outsiders were unable to resolve the debate because they were uncertain about who was supposed to take the lead. For example, one district staff developer stated, "When there was a problem that needed to be solved, no one knew who was in charge. We didn't know if it was the [external] people in charge or who it was that was in charge of the whole thing." And similarly, an outside consultant explained,

So, this [approach to partnership] is very new to me. And I think that's why I'm very, very tentative. I'm very unsure of myself. I'm very worried about offending people. And at the same point in time I'm very concerned about people going off in directions that I as a professional . . . feel are ill-advised. And yet being very unsure of when to step in and say "That just is really not sound."

In this instance, normative agreement about authority relations between district staff and external consultants failed to emerge.

As predicted by Dornbusch and Scott (1975), it was very difficult to engage in joint work in the absence of clear authority relations. Work on the literacy institutes stalled as those in the district and the external consultants could not resolve their very different ideas about next steps for the summer professional development. Ultimately, the senior member of the research center—who had informal authority—stepped in, authorized the district professional development staff to take the lead, and scaled back the participation of external partners. From that point on, the external partners moved into a more traditional consultant role, providing feedback at the request of the district professional development staff and stepping in to give a talk during one segment of the summer professional development. The conflict was diffused and tensions eased. Authority relations had been established, in this case with district staff developers in charge.

Authority is inherently relational and therefore contextual. That is, an individual has authority only in relation to others. Thus, one could have authority in one configuration of participants but have little authority in another configuration, even within the same instance of collaboration. For example, the external partner who was authorized by the superintendent to lead the summer institutes had a great deal of authority when he worked with some division directors and mid-level administrators like the district professional development providers. At the same time, this person had little authority when interacting with the executive leadership in the district on the literacy institutes. Similarly, the same individual had shared authority with the director of the district research division as they worked together to identify demonstration teachers but had little authority with the assistant superintendent in the leadership development work.

Finally, authority relations were not stable. Rather, they were likely to evolve through interaction and social negotiation (see also Pace & Hemmings, 2007, on this point). In fact, over the course of the initiative, authority relations gradually evolved such that insiders had greater authority than outsiders in collaborative groups. In the first year of the initiative, outside partners had greater authority than insiders in aspects of four of the seven instances of collaboration. By the end of the second year of the initiative, authority relations evolved or were explicitly renegotiated such that district personnel had greater authority in insider–outsider collaborative groups in all but one instance of collaboration.

### *Status*

Prior scholarship on insider–outsider relationships has argued that outsiders are frequently accorded status in collaborations, especially if they are academics (Bickel & Hattrup, 1995; Goodlad & Sirotnik, 1988; Osajima, 1989; Sinclair & Harrison, 1988). For example, Osajima contended that school personnel have historically been situated in a subordinate position in insider–outsider collaborations

because scientific theory and knowledge is privileged over the practical knowledge of school people. However, we found a more complicated scenario than this. Individuals often granted status to others when they perceived them to have special skills or attributes, particularly knowledge of research, knowledge of practice, or practical experience. Although some individuals did grant others status because of their academic credentials or specialized training, both insiders and outsiders were more likely to attribute status to others because of practical experience than research knowledge or academic credentials. Furthermore, attributions of status to an individual were not comprehensive but rather were quite specific. That is, they were accorded to an individual for a particular domain of work but not for others.

There were multiple criteria for attributing status among both the insiders and outsiders involved in the partnership. Some insiders and outsiders did in fact grant others status for their academic knowledge or credentials, as suggested by prior scholarship. For example, a member of the research center mentioned that she saw two academic researchers as highly valued members of the collaboration because of what she perceived to be deep knowledge about the research literature. She reports thinking to herself, "Okay, good. Two people that really know the content. I mean, I really asked them a lot of questions when I first met them, and really felt like these are two very good people, they're highly qualified to do this, and they would be good on a team." However, many insiders and outsiders also granted or denied status to individuals based on their practice-based knowledge. For example, one district employee denied the status of another colleague in the district, saying,

I sat in some of those preplanning meetings across from a psychologist who never taught a day in that person's life in the classroom, telling me how to teach and what's important for first graders. I'm like screw that. Completely. Because you don't know. You've read a lot of stuff, but you don't know that when I have 25 first graders in my room, there's five different ways I need to teach reading and you think that because you read this research, and you revere it to be whatever, that that's the way I'm supposed to teach my kids to read? I don't think so.

Furthermore, insiders and outsiders granted status based on an individual's clinical knowledge, particularly experience in urban schools serving ethnically and linguistically diverse children. For example, several insiders accorded status to an outside academic because "he also is still very much involved in a school setting that has some of the same demographics that we'd see in the schools here in [the district]."

Contrary to the findings from prior scholarship, insiders and outsiders alike were more likely to grant others status based on having practical experience than having research knowledge or an academic credential. As can be seen in Table 2, 64% of those interviewed accorded status to others in the partnership based on their teaching experience or experience providing professional development, whereas



TABLE 2  
Criteria for Granting Others Status

	Insiders	Outsiders	Total
Credential/Academic training	24%	8%	18%
Academic knowledge/Knowledge of research	66%	50%	60%
Practical experience/Practitioner knowledge	57%	75%	64%
Specialized knowledge of urban schools, teachers, or children	29%	8%	21%
Other	33%	33%	33%

*Note.*  $N = 21$  insiders; 12 outsiders.

60% of individuals accorded status based on knowledge of research and 18% accorded status on the basis of an academic credential. Furthermore, outsiders were actually more likely to accord status based on practical or professional experience than insiders. Seventy-five percent of outsiders accorded status to others based on their practical experience, whereas only 57% of district personnel did. As noted earlier, individuals did not make blanket attributions of status to others. Rather, status attributions were conditional on particular dimensions and came into play only when the joint work touched on those dimensions. For example, the outsider involved in the difficult negotiation related to the literacy work was accorded status for her expertise in reading instruction by some of the same individuals who denied her status in professional development. Similarly, a key leader of mathematics in the district was accorded status for her content knowledge in mathematics but was disparaged for her lack of experience teaching the particular curriculum at the heart of the summer professional development.

Ultimately, both insiders and outsiders were more likely to grant status to outsiders for both research knowledge *and* practical experience. Thus, 12 outsiders involved in the collaboration were granted status by others based on knowledge of the research (all 12 were accorded status by insiders, and 3 of those 12 were also accorded status by outsiders). In contrast only 2 insiders were accorded status based on their knowledge of research, all by outsiders. Nine outsiders were accorded status for their practical experience as classroom teachers or professional development providers (8 were accorded status by insiders and 1 by outsiders). In contrast 7 insiders and the staff of two divisions were accorded status for their practical experience as classroom teachers or professional development providers (5 individuals and one division by outsiders and 2 individuals and one division by insiders). Insiders were also much more likely to receive negative attributions based on lack of practical experience than outsiders. Thus, 8 individuals in the district and five entire divisions were disparaged for their lack of practical experience (mainly by insiders, but outsiders also critiqued four individuals). In contrast, and

perhaps surprisingly, only two outsiders were disparaged for their lack of practical experience.

Although analytically distinct, attributions of status were, at times, related to authority relations. Those with formal authority were more likely to authorize someone—granting them informal authority—if they accorded status to that person because of his or her expertise. In all but one instance where an insider or outsider was granted informal authority to lead an aspect of the work, this individual was also granted status by the person doing the authorizing. At the same time, in the three instances where authority relations failed to emerge through negotiations between insiders and outsiders, insiders did not grant outsiders status and outsiders did not accord insiders status. For example, in the case of the literacy institute described earlier, both insiders and outsiders saw themselves as having expertise in professional development and neither saw the other as being particularly knowledgeable or experienced in it. As neither side granted status to the other, neither side endorsed the authority of the other. Progress stalled until a senior researcher with informal authority authorized the district staff developers to take the lead in the joint work. However, in spite of these links between status and authority, in nearly all instances of collaboration, there were individuals with authority and no status, and also individuals with status but no authority.

### INFLUENCE OF AUTHORITY RELATIONS AND STATUS ON FRAME DYNAMICS

Authority relations and attributions of status were consequential because they shaped the process by which insiders and outsiders negotiated the direction of their joint work. In nearly all instances of collaboration, there were times when there were differences of opinions about the best course of action. When this happened, individuals put forth ideas about particular goals and suggested tactics for achieving those goals. In the language of frame theorists, they engaged in prognostic framing (Snow et al., 1986; R. H. Williams & Kubal, 1999). In so doing, they made arguments to one another, drawing on research, previous experience, and the facts on the ground in an attempt to persuade others of the direction to go. Ultimately, groups were only able to move forward in their planning once a given proposal began to achieve what frame theorists call resonance. That is, once a particular solution frame began to “make sense” to others in the group, it generated momentum and the work was able to progress in a particular direction.

We found that those with formal or informal authority had different tools available to them to bring to the persuasive process than did those with limited authority in the collaborative group. Status also played a role, but it was less influential and, at times, mediated by authority relations. Finally, individuals chose to use different tools in the effort to persuade, and those tools had different degrees

of success when collaborators held diverging views versus when the views were more homogeneous. Here, we illustrate these claims by describing the tactics individuals who were differentially positioned in collaboration used to argue for their position. We also evaluate the degree to which these tactics mobilized others and shaped the direction of the work.

### *Authority and Status*

A small number of individuals involved in collaborative efforts as part of PDR had authority (formal or informal) and were granted status by others in their collaborative partnerships. Mostly, those with both authority and status were insiders, but there were at least two key outsiders who had been granted informal authority and status. Those with both authority and status tended to rely on persuasion to influence the direction of the collaboration. That is, they put forth ideas about appropriate solutions and backed them up with arguments that drew on their analysis of the nature of the district; their own prior experience; or, at times, references to research. Those with both status and authority were remarkably successful in their framing activities. We judged success by the degree to which others in a group took up and argued for a given position as their own (a key indicator of the resonance of a frame) or the degree to which the frame shifted the central terms or direction of the debate.<sup>5</sup>

For example, in the initial discussions of the overall design of the professional development, outsiders argued for a strategy of depth suggesting that the district could make use of its resources by focusing more intensively on a subset of its schools. However, the superintendent—who had formal authority but also was accorded status by insiders and outsiders alike—offered a counterframe, arguing that they should include all schools in the professional development initiative. She justified this approach by drawing on recent research on the importance of systemic approaches to instructional improvement and argued that what the district really needed was a uniform approach to instruction. She contended that including all schools in the initiative would best foster a uniform approach to instruction that would meet the needs of the districts' highly mobile student population. This argument was persuasive to both insiders and outsiders involved in this decision point, who were generally familiar with and supportive of the notion of systemic reform. As an indicator of the resonance of this argument, insiders and outsiders alike repeated this logic to one another in subsequent conversations. Ultimately, the summer professional development institutes were designed to include teams from every school in the district. In all but one decision point that involved individuals with both status and authority, prognostic frames put forth by these

---

<sup>5</sup>For more details on the dynamics of resonance, see Coburn (2006).

individuals generated resonance with others, shaping the direction of the joint work in consequential ways.

One reason for the success of their framing activities is that individuals with status and authority were able to use their authority to create conditions that would make their frames more likely to resonate. We identified two key strategies that those with both authority and status used. First, they frequently set the agenda for discussions in ways that privileged approaches that they were advocating. For example, in the case of collaboration around the mathematics institute, the mathematics leadership favored including Lesson Study as a central part of the professional development strategy in the second year. Although there were many ideas for ways to continue the work started in the first year, the mathematics leadership put discussion of this approach on the agenda of a key planning meeting. The team discussed the approach and decided to include it. Although the mathematics leader did not in any way compel the mathematics team to embrace the approach, she did privilege the approach by putting it and not other approaches on the agenda. In this way, she used her authority to play an influential role in the ongoing debate about the future directions of math work.

Second, those with authority and status influenced frame debates by controlling who participated in the discussion, often inviting those who were like-minded to participate, a tactic we call narrowing participation. For example, in the face of controversy in the district about appropriate ways to teach mathematics, the district mathematics leadership, who favored constructivist approaches to mathematics instruction, sought out other district personnel and outside consultants who were knowledgeable about and committed to using the constructivist curriculum that the district had adopted to participate in design work. In so doing, the mathematics leadership created a team of insiders and outsiders that had remarkably similar points of view about what constituted good mathematics instruction. This created a very different context within which frame dynamics unfolded than in the design of the literacy institute, which involved representatives from the many diverse views about good literacy instruction inside and outside the district.

Those with status and authority used agenda setting or narrowing participation in 40% of the decision points in which they were involved. But, it is also important to note that those with authority and status were also quite successful in the absence of these tactics, suggesting that their individual credibility and skill at framing also supported their influence in negotiations with their partners.

### *Authority Only*

Some individuals involved in the collaboration had authority, but not status. This was most common with upper-level administrators who had a great deal of



authority but did not have credibility with either the outsiders or those below them on matters of instruction and professional development that were at the heart of PDR. Those with authority but not status were much more likely to use tactics such as agenda setting or narrowing participation to support prognostic framing than did those with both status and authority. They used these tactics in more than two thirds of the decision points in which they participated. Again, these tactics were almost always successful, as frames put forth under these conditions were more likely to generate resonance with others involved in the deliberation about future directions.

However, individuals with authority but no status also used their authority directly to influence the direction of the negotiation, without attempting to persuade or in other ways bring along others. We saw several instances when those in authority rejected, overturned, or stalled the work done in partnerships at lower levels in the district. For example, in the case of leadership development work, members of the outside research center worked with the principal's leadership council to develop plans for systemwide professional development for school leaders. This group put forth a plan that was substantially different from what the district currently did and, as it turns out, from ideas about good principal professional development held by a key assistant superintendent. This assistant superintendent, in turn, never responded to the work publicly in the context of a planning meeting, although she criticized it privately to a few people. The plan was never acted upon and the work stalled.

We also saw instances where individuals with authority but not status compelled those below them or outside partners to take the work in a particular direction. For example, one of the assistant superintendents who favored adopting a behaviorist math program as a supplement to the district's constructivist math textbook insisted that the math team incorporate the supplemental program into the summer institute work and the follow-up professional development. As one of the math staff developers explained, "The fifth grade has an additional piece in that politically we have this [supplemental program] issue. And because that is something that is coming down from the top, and we're being scrutinized, we had to build that in to this follow-up as well." In these instances, the direction of the work was altered not because those in the partnership had been persuaded that it was the most appropriate route to go, orchestrating what Binder (2002) called an ideological shift. Rather, the direction of the partnership was shaped because partners were compelled, as those in authority orchestrated what Binder called a political shift. It is important to note that those with authority were most likely to engage in this set of tactics when they had quite different beliefs about the appropriate way for the initiative to go than others in the collaborative group. Ultimately, individuals who used their authority in this way were quite successful in shaping the direction of the collaborative work; direct uses of authority were successful

in shaping the direction of the work in every decision point where they were used.

### *Status Only*

There were numerous individuals involved in PDR who had status but not authority. Most of these individuals were outsiders who had limited authority either through the terms of their contract or as the result of an emergent negotiation of authority relations that granted insiders greater authority. Like those with status and authority, those with status alone relied primarily on persuasion to influence the direction of collaboration. However, unlike those with status and authority, those with status alone were not nearly as successful in persuading others of the direction they thought the work should go. They were only successful in generating resonance for their frames in half of the decision points in which they were involved. Furthermore, they were most successful when they were engaged in collaboration with others who had similar beliefs about what constitutes good instruction or high-quality professional development. In other instances where there were more diverse views about the appropriate direction to go, those with status alone were less able to frame ideas in ways that generated resonance with those with greater authority in the group, although it did happen from time to time.

### *Neither Status Nor Authority*

Finally, some insiders and outsiders had neither status nor authority in collaborative partnerships. In this case, attempts to persuade others were uniformly unsuccessful. This phenomenon is perhaps best illustrated by an outsider whose status was denied in the first year by insiders but who then came to be seen as quite knowledgeable by some of the same insiders in the second year. This outsider made many of the same arguments for the direction of the literacy work in both years, but the arguments were rejected by insiders in the first year when the outsider lacked status. She was subsequently influential in the direction of the work in the second year once she came to be seen as having expertise, providing evidence for the important role that credibility plays in the success of framing activities.

In the absence of success using persuasion, some individuals with neither status nor authority resorted to other tactics to influence the direction of the collaborative work. The first strategy these individuals used was to have others who they saw as having status or authority promote their ideas. This happened at four decision points. For example, in the second year of the initiative, there was controversy about whether a particular approach to reading instruction was appropriate for the

poor students and students of color that the district served. One African American member of the planning team explained,

Personally, one of my feelings is that if . . . we don't start to look at the cultural piece—in that if I'm different than these children and I don't respect what they bring to the table—then [this program] is not going to address that.

A key advocate for the instructional approach who was White responded to this criticism by bringing in an African American academic she knew to advocate the approach she favored in order to help develop credibility for her argument. This academic explained why he was brought into the work in the following way:

It's very difficult to sometimes be a prophet in your own land. Like, [the district insider] had a concern about "they've been hearing from me and hearing from me and hearing from me, and we need for someone to parrot what I say, but a different face—and you do just that, you and I are in synch, and they're going to listen to you."

This tactic, which we call the use of frame articulators (Turner, 2008), was successful at all four decision points where it was used. In this instance, those who were initially skeptical raved about the instructional approach and the outside academic, pointing to both his academic expertise and his experience as an African American as contributing factors to his ability to be helpful. One former skeptic said, "So he was helpful there. Just, he is African American so he has personal experiences to draw from as well." The controversial approach to literacy instruction became the centerpiece of this team's professional development in the second year of the initiative.

The second strategy employed by those without status or authority was to enlist others with authority to intervene on their behalf. Two insiders used this strategy at separate decision points. For example, when the literacy team was involved in a dispute about which approaches to promote during the second year of their professional development initiative, one member of the literacy team got an assistant superintendent to intervene to mandate the approach that this member of the literacy team favored. In both instances, this tactic was successful as the intervention from those with authority shaped the direction of the work.

Finally, one outsider sought to gain legitimacy for his approach by invoking the authority of others—in this case, the authority of the foundation that was supporting the initiative. Unlike the other two strategies for gaining leverage for those without status or authority, this strategy backfired. Those in the district saw it as a breach, and it, along with several other incidents, prompted a call for clarification of authority relations and an explicit renegotiation of appropriate roles for insiders and outsiders in the second year of the initiative. As was the

case with those with authority but no status, those with no status and no authority were most likely to attempt these power tactics when they were negotiating with others who had substantially different ideas about the direction to proceed in the collaborative group.

All of this suggests that the tactics used in frame debates were greatly influenced by an individual's position in relation to others in the group. Authority relations were particularly influential, considerably more so than status relations. Those with authority had a much greater range of strategies for influencing the direction of collaborative work and felt free to use them. Those with authority were more successful than those without authority when using tactics available to both groups. For example, those with authority were two times more likely to persuade others on the basis of prognostic framing alone than those with status but no authority. Those with status were most able to be influential to the degree that they gained informal authority by being authorized from above or endorsed from below. In this way, authority relations often mediated the influence of status in negotiations between insiders and outsiders.

This analysis also suggests that the use and success of framing tactics depends upon the diversity of beliefs in a given group. Persuasion was less likely to be successful—even for those with authority or status—when there were divergent views about the appropriate direction to go. In fact, those with authority often used their position to limit negotiation to those who shared their point of view, a tactic that enabled greater influence in the deliberation process. Similarly, individuals were much more likely to use—or get others to use—direct control strategies in the face of diverse views. This suggests that authority relations are even more important in determining the direction of collaborative work when there is a diversity of beliefs in a given partnership.

## INSIDER–OUTSIDER PARTNERSHIPS AT THE DISTRICT LEVEL

Most research on insider–outsider partnerships investigates the phenomenon at the school level or in out-of-school settings. Yet the district central office level is considerably more complicated, politicized, and fluid than a school setting. Many school districts have highly complex and departmentalized organizational structures (Hannaway, 1989; Meyer & Scott, 1983; Rowan, 1986; Spillane, 1998). There are multiple levels of the system from the executive level, to directors of divisions, to frontline administrators who are often charged with carrying out the details of the work. There are also multiple divisions that are implicated in matters of instruction, typically including curriculum and instruction divisions, assessment and testing divisions, and special education divisions, to name a few (Spillane, 1998). Furthermore, organizational structure and authority relations are often fluid



as changes in upper level administrators lead to reorganization and shifting roles. This complex and fluid structure influenced the role of status and authority in insider–outsider negotiations in at least four ways.

First, in this district as in many districts, there were uncertain authority relations among different levels and divisions of the district. The main lines of authority went from the superintendent to the assistant superintendent in charge of each level of schooling (one each for elementary, middle, and high) to school principals. But there were uncertain authority relations between these assistant superintendents and the heads of the main divisions involved in instruction, including Curriculum and Instruction, Assessment, a division in charge of English Language Learners, and Special Education. In the second year of the initiative, the district appointed a chief academic officer with responsibility over all these divisions, but the relationship between the assistant superintendents and the instruction divisions remained ambiguous. Uncertain authority relations *within* the district led to complications for insider–outsider partnerships that stretched across the multiple divisions. For example, in the second year of the initiative, outsiders worked with the mathematics division to identify an outside provider who could provide professional development to school leaders in high quality mathematics instruction. The outsider conferred with the chief academic officer—the supervisor of the Curriculum and Instruction division—to gain approval of the plan but did not confer with the assistant superintendents. The assistant superintendents, in turn, saw professional development for school leaders as under their purview. Thus they viewed this move as an affront to their authority and saw the outsider as out of line. Thus, it is not just authority relations between insiders and outsiders that influenced insider–outsider partnerships but also authority relations *within* the district itself. These authority relations were made more complicated by the complexity of the district central office.

Second, the organizational structure of the school district also shaped negotiation because of the rather loose linkages between different levels of the district hierarchy. In all but two instances of collaboration, outsiders worked simultaneously with individuals at multiple levels of the system to negotiate the direction of the work. Negotiations between outsiders and top-level administrators led to broad directions for the work. The details of implementation were then developed in negotiation with frontline administrators who were responsible for carrying out the work. In the absence of tight linkages between the top and bottom of the system, there was often a somewhat tenuous relationship with the collaborative decisions made at the top and those made by inside–outside partners at the bottom of the system. For example, in the first year of the initiative, the superintendent made it clear that she wanted attention to issues of diversity to sit at the center of summer institutes in reading and mathematics. Yet because it was not a priority for frontline administrators and because the frontline administrators either had informal authority over outsiders (in mathematics) or there were contentious

authority relations (in literacy), the resulting design paid only symbolic attention to issues of diversity. It was difficult for outside partners to coordinate between levels of the system in the absence of mechanisms within the district to achieve that coordination themselves.

There was another outcome of the multilevel design process just described. In five of the seven instances of collaboration, outside partners worked most closely and in most detail with frontline administrators. During the course of that joint work, there was a process of mutual influence whereby outsiders persuaded insiders and, in some instances, insiders persuaded outsiders of particular directions to go. However, executive-level decision makers rarely took part in this level of conversation. Thus, they did not have the opportunity to participate in the frame debates or be persuaded by them over time. For this reason, executive-level decision makers who had formal authority over the proceedings were particularly likely to reject the work done by those at lower levels of the system or insert things into the process that were not in line with the ongoing direction of the conversation at lower levels of the system. We saw this phenomenon in three of the seven instances.

Finally, turnover is endemic at the upper levels of school districts, and this district was no exception. During the three years covered by our study, the district lost its longtime superintendent, had an interim for a year, and then at the end of our study hired a new superintendent. Turnover at the top of the system had a ripple effect on the authority relations guiding the negotiation. Those with informal authority were particularly vulnerable. For example, the first superintendent authorized several outsiders to take the lead on key aspects of the initiative. These outsiders, in turn, were endorsed by others in the district. But once the superintendent left and a few of the key positions in the next layer of the district leadership changed, this history of authorization and endorsement was lost. Thus, when these outside individuals took the lead in a manner consistent with the established authority relations under the prior superintendent, new members of the district leadership viewed the outsiders as overstepping their role. Normative agreement about appropriate roles for the outsider that was forged under the original superintendent began to unravel with the presence of new people with formal authority. Ultimately, this was not resolved until there was an explicit renegotiation of the role for outsiders throughout the initiative, which resulted in shifting authority more firmly to the insiders, especially at the executive level.

## DISCUSSION AND IMPLICATIONS

As districts seek to create instructional improvement at scale, they are increasingly reaching out to external organizations to assist them with this endeavor. Yet the

potential of these relationships for bringing about instructional improvement is related to not only the quality of the advice or assistance these organizations offer but also the nature and dynamics of the relationship that outsiders and insiders are able to forge with one another. Our analysis suggests that status and authority relations play a key role in shaping the nature of these relationships. Authority relations are particularly important because the absence of normative agreement about authority can lead to conflict, misunderstandings, and an inability to move the work forward. But authority relations are also important because they shape how negotiation unfolds. Those with authority are privileged in the social negotiation about directions for the partnership. They have a greater range of tools for persuasion at their disposal and the ability to use more direct mechanisms of control to impact the direction of the partnership. Attributions of status are also important, but often less so than authority. If outsiders or insiders have status but not authority, they must rely on their personal credibility and the wisdom of their arguments to persuade those who have authority to move in particular directions.

All of this is more complicated and challenging when negotiation unfolds in the context of a district central office. The multileveled structure of school districts, combined with their uncertain authority relations and loose connections between levels, makes it more difficult to forge and maintain normative agreement about authority. Endemic turnover requires that authority relations and status hierarchies be negotiated repeatedly as new individuals become involved in partnership activities, creating new expectations for roles and relationships and new and sometimes different attributions of status.

These findings have several implications for our understanding of insider-outsider relationships. First, this research highlights the importance of careful attention to authority relations. Those involved in crafting partnerships may shy away from explicit attention to authority because it seems contrary to democratic ideals embedded in the notion of collaboration.<sup>6</sup> Although it may seem counterintuitive to some, this study suggests that the development of clear authority relations actually enables productive working relationships. Shared understanding of appropriate roles and relationships provides guidance for interaction and decision making, and for mitigating against breaches, power struggles, and misunderstanding. In fact, as suggested by Dornbusch and Scott (1975) and illustrated by this study, in the absence of clear authority relations, it can be very difficult to move forward.

This finding has implications for partnerships like PDR that seek to craft alternative forms of relationships between insiders and outsiders. The leaders of PDR intended to create a new kind of partnership, with the goal of maximizing the rich knowledge that researchers, experienced practitioners and professional development providers outside of the district, as well as individuals at multiple

---

<sup>6</sup>See Pace and Hemmings (2007) on this point related to classroom authority.



levels inside the district have to offer. The approach to co-construction guiding the partnership was supposed to mitigate against status issues that privilege academic knowledge and to create shared stakes and shared decision making. Yet the fact that this form of partnership was so unfamiliar to both insiders and outsiders involved in the work probably contributed to some of the difficulties in establishing normative agreement about appropriate authority relations. In addition, although the partnership strove to create a sense of shared stakes, it can be argued that the district personnel had much more to gain and lose than the external consultants. Authority relations were most clearly defined—and normative agreement was easiest to develop and sustain—when the terms of the partnership resembled a traditional consulting role. Relations were most likely to be bumpy and conflictual when authority relations developed emergently and when insiders and outsiders had different ideas about what role each was supposed to play. This suggests that initiatives that seek nontraditional partnerships must work extra hard to develop and communicate clear models of authority relations and clear expectations about what it means for insiders and outsiders to work with one another in this fashion.

Second, this study raises questions about prior scholarship on the importance of status in influencing insider–outsider relationships. It suggests that status relations are less unitary, more situational, and more fluid than that portrayed in prior scholarship. Furthermore, it suggests that there may be multiple criteria for attributing status operating simultaneously. Individuals, both in districts and outside of districts, grant status not only to outsiders who they perceive to have research knowledge or expertise but also to those who they perceive to have great practical knowledge or extensive experience. It is possible that the large percentage of outsiders who granted status based on perceptions of practical experience in this project was related to the fact that the leaders of PDR promoted the value of knowledge from practice so strongly. But this does not explain why so many insiders were more likely to grant status to others on the basis of practical experience than on the basis of research knowledge. This suggests that rather than *assuming* that the privileging of academic knowledge over practical knowledge in the larger environment influences how individuals make status attributions in the context of a local partnership, it is important to *investigate* the nature of status attribution directly.

Third, this article contributes to the scholarship on insider–outsider partnerships by highlighting the role of organizational structure in the dynamics of negotiation and, ultimately, in how partnerships unfold. Investigating the dynamics of framing at the district central office level brings the role of organizational structure into relief. It suggests that organizational structure shapes negotiation in part because of the way it structures authority relations. Individuals in different positions of the district hierarchy are accorded different levels of formal and informal authority. Outsiders, even those who are accorded informal authority, are differently positioned depending on whether they are interacting with those at the top or those at



the bottom of the system. These authority relations, in turn, influence the dynamics of negotiation in substantial ways.

The connections between different areas of the district are consequential for negotiation. Uncertain authority relationships between district divisions create ambiguity and increase the opportunities for missteps and breaches. Loose connections between the bottom and the top of the system create great challenges for communication and coordination. Thus, although partnerships at the district central office level may seek to help districts solve some of the vexing organizational challenges that appear to impede their ability to foster instructional improvement at scale, the partnerships are subject to some of the same organizational dynamics themselves.

Fourth, this analysis also has implications for attempts to use collaborative partnerships to leverage district change. More specifically, it suggests that outsiders are most likely to be able to leverage change in the district when they have similar points of view as those on the inside. In this study, insiders were more likely to have formal or informal authority in collaborative groups, especially as the partnership evolved over time. Outsiders were more likely to have status. Given that authority was much more influential than status, outsiders (and some insiders) found themselves mainly relying on their ability to persuade those with authority of the wisdom of their approach. However, persuasion was less likely to be successful under conditions of diverse views about appropriate directions for a particular initiative. This suggests that in the absence of shared beliefs about the direction for the collaborative work, those with status may face considerable difficulty if they attempt to promote approaches that diverge substantially from those approaches that are valued by those in positions of authority in the district. Indeed, in PDR, outsiders with status but no authority were most likely to be successful in shaping the direction of district work when they had shared understandings with at least a subset of insiders with whom they collaborated. Attempts to promote directions for the initiative that departed substantially from what those with authority were familiar with and believed in were frequently unsuccessful, as was the case with the leadership development work proposed by outsiders. This suggests the promise of a more incremental, long-term approach to systemic change than is typically sought, at least for collaborative partnerships operating under this set of authority relations.

Finally, this work suggests the benefit of future research on partnerships with different configurations of authority. This investigation provides insight into the development and importance of authority and status relations. But it raises the questions: Will insiders be more likely to have authority and outsiders be more likely to have status under different strategies for establishing partnerships? Will authority relations be as challenging to establish and maintain as with this approach? It is only through continued investigation of the nature and role of authority that we will begin to better understand how different configurations of

authority relations and status create conditions that are more or less conducive for partnerships to support district improvement over time.

## REFERENCES

- Balkwell, J. W. (1994). Status. In M. Foschi & E. J. Lawler (Eds.), *Group processes: Sociological analyses* (pp. 119–148). Chicago: Nelson-Hall.
- Barley, S. R. (1990). Images of imaging: Notes on doing longitudinal field work. *Organizational Science*, 1(3), 220–247.
- Benford, R. D., & Snow, D. A. (2000). Framing processes and social movements: An overview and assessment. *Annual Review of Sociology*, 26, 611–639.
- Bickel, W. E., & Hatrup, R. A. (1995). Teachers and researchers in collaboration: Reflections on the process. *American Educational Research Journal*, 32(1), 35–62.
- Binder, A. J. (2002). *Contentious curricula: Afrocentrism and creationism in American public schools*. Princeton, NJ: Princeton University Press.
- Boostrom, R., Jackson, P. W., & Hansen, D. T. (1993). Coming together and staying apart: How a group of teachers and researchers sought to bridge the ‘research/practice’ gap. *Teachers College Record*, 95(1), 35–44.
- Brookhart, S. M., & Loadman, W. E. (1992). School-university collaboration: Across cultures. *Teaching Education*, 4(2), 53–68.
- Bryk, A., Rollow, S. G. & Pinnell, G. S. (1996). Urban school development: Literacy as a lever for change. *Educational Policy*, 10(2), 172–201.
- Burch, P. E. (2002). Constraints and opportunities in changing policy environment: Intermediary organizations’ response to complex district contexts. In A. M. Hightower, M. S. Knapp, J. A. Marsh, & M. W. McLaughlin (Eds.), *School districts and instructional renewal* (pp. 111–126). New York: Teachers College Press.
- Coburn, C. E. (2006). Framing the problem of reading instruction: Using frame analysis to uncover the microprocesses of policy implementation in schools. *American Educational Research Journal*, 43(3), 343–379.
- Coburn, C. E., Honig, M. I., & Stein, M. K. (in press). What is the evidence on districts’ use of evidence? In J. Bransford, L. Gomez, D. Lam, & N. Vye (Eds.), *Research and practice: Towards a reconciliation*. Cambridge, MA: Harvard Educational Press.
- Coburn, C. E., Toure, J., & Yamashita, M. (2009). Evidence, interpretation, and persuasion: Instructional decision making in the district central office. *Teachers College Record*, 111(4).
- Cohen, E. G. (1994). *Designing groupwork: Strategies for the heterogeneous classroom. Second edition*. New York: Teachers College Press.
- Corcoran, T. B., & Rouk, U. (1985). *Using natural channels for school improvement: A report on four years of the urban development program*. Philadelphia: Research for Better Schools, Inc.
- Cress, D. M., & Snow, D. A. (2000). The outcomes of homeless mobilization: The influence of organization, disruption, political mediation, and framing. *American Journal of Sociology*, 105(4), 1063–1104.
- Davies, S. (1999). From moral duty to cultural rights: A case study of political framing in education. *Sociology of Education*, 72(1), 1–21.
- Dornbusch, S. M., & Scott, W.R. (1975). *Evaluation and the exercise of authority*. San Francisco: Jossey-Bass.
- Erickson, F., & Christman, J. B. (1996). Taking stock/making change: Stories of collaboration in local school reform. *Theory into Practice*, 35(3), 149–157.
- Feldman, M. A. (1995). *Strategies for interpreting qualitative data*. Thousand Oaks, CA: Sage.

- Firestone, W. A., & Fidler, J. L. (2002). Politics, community, and leadership in a school-university partnership. *Educational Administration Quarterly*, 38(4), 449–493.
- Fligstein, N. (2001). Social skill and the theory of fields. *Sociological Theory*, 19(2), 105–125.
- Freedman, R., & Salmon, D. (2001). The dialectic nature of research collaborations: The relational literacy curriculum. In T. Ravid & M. G. Handler (Eds.), *The many faces of school-university collaboration: Characteristics of successful partnerships* (pp. 179–183). Englewood, CO: Teachers Ideas Press.
- Gamoran, A., Anderson, C. W., Quiroz, P. A., Secada, W. G., Williams, T., & Ashmann, S. (2003). *Transforming teaching in math and science: How schools and districts can support change*. New York: Teachers College Press.
- Gamson, W. A. (1992). *Talking politics*. Cambridge, UK: Cambridge University Press.
- Gates-Duffield, P., & Stark, C. (2001). Recipes for professional development schools: A metaphor for collaboration. In T. Ravid & M. G. Handler (Eds.), *The many faces of school-university collaboration: Characteristics of successful partnerships* (pp. 45–55). Englewood, CO: Teachers Ideas Press.
- Gifford, B. R. (1986). The evolution of the school-university partnership for educational renewal. *Education and Urban Society*, 19(1), 77–106.
- Glennan, T. K., & Resnick, L. B. (2004). School districts as learning organizations: A strategy for scaling education reform. In T. K. Glennan, S. J. Bodilly, J. R. Galegher, & K. A. Kerr (Eds.), *Expanding the reach of education reforms: Perspectives from leaders in the scale-up of educational interventions* (pp. 517–564). Santa Monica, CA: RAND Education.
- Goldring, E., & Sims, P. (2005). Modeling creative and courageous school leadership through district-community-university partnerships. *Educational Policy*, 19(1), 223–249.
- Goodlad, J. I., & Sirotnik, K. A. (1988). *School-university partnerships in action: Concepts, cases and concerns*. New York: Teachers College Press.
- Handler, M. G., & Ravid, R. (2001). *The many faces of school-university collaboration: Characteristics of successful partnerships*. Englewood, CO: Teachers Ideas Press.
- Hannaway, J. (1989). *Managers managing: The workings of an administrative system*. New York: Oxford University Press.
- Hasslen, R., Bacharach, N., Rotto, J., & Fribley, J. (2001). Learning connections: Working toward a shared vision. In T. Ravid & M. G. Handler (Eds.), *The many faces of school-university collaboration: Characteristics of successful partnerships* (pp. 60–72). Englewood, CO: Teachers Ideas Press.
- Heckman, P. E. (1998). The Southern California Partnership: A retrospective analysis. In K. A. Sirotnik & J. I. Goodlad (Eds.), *School-university partnerships in action: Concepts, cases and concerns* (pp. 106–123). New York: Teachers College Press.
- Honig, M. I. (2004). The new middle management: Intermediary organizations in education policy implementation. *Educational Evaluation and Policy Analysis*, 26(1), 65–87.
- Isabella, L. A. (1990). Evolving interpretations as a change unfolds: How managers construe key organizational events. *Academy of Management Journal*, 33(1), 7–41.
- Keating, P. J., & Clark, R. W. (1988). Accent on leadership: The Puget Sound Educational Consortium. In K. A. Sirotnik & J. I. Goodlad (Eds.), *School-university partnerships in action: Concepts, cases and concerns* (pp. 148–166). New York: Teachers College Press.
- Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, G. (2006). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education*, 112(4), 496–520.
- Kornfield, J., & Leyden, G. (2001). Working together: A successful one-to-one collaboration. In T. Ravid & M. G. Handler (Eds.), *The many faces of school-university collaboration: Characteristics of successful partnerships* (pp. 194–206). Englewood, CO: Teachers Ideas Press.

- Lieberman, A. (1988). The Metropolitan School Study Council: A living history. In K. A. Sirotnik & J. I. Goodlad (Eds.), *School-university partnerships in action: Concepts, cases and concerns* (pp. 69–86). New York: Teachers College Press.
- Marsh, J. A., Kerr, K. A., Ikemoto, G. S., Darilek, H., Suttrop, M., Zimmer, R., et al. (2005). *The role of districts in fostering instructional improvement: Lessons from three urban districts partnered with the Institute for Learning*. Santa Monica, CA: RAND Corporation.
- McAdam, D., & Scott, W. R. (2005). Organizations and movements. In G. F. Davis, D. McAdam, W. R. Scott, & M. N. Zald (Eds.), *Social movements and organizational theory* (pp. 4–40). Cambridge, England: Cambridge University Press.
- Meyer, J. W., & Scott, W. R. (1983). *Organizational environments: Ritual and rationality*. Beverly Hills, CA: Sage.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.
- Osajima, K. H. (1989). Building school-university partnerships: Subjectivity, power, and the process of change. *The Urban Review*, 21(2), 111–125.
- Pace, J. L., & Hemmings, A. (2007). Understanding authority in classrooms: A review of theory, ideology, and research. *Review of Educational Research*, 77(1), 4–27.
- Ravid, T., & Handler, M. G. (2001). *The many faces of school-university collaboration: Characteristics of successful partnerships*. Englewood, CO: Teachers Ideas Press.
- Rosen, L. (2008). *The teaching school : Reinventing the role of research universities in K-12 education*. Unpublished manuscript, Center for School Improvement, Chicago.
- Rowan, B. (1986). *Rationality and reality in instructional management: Results from a survey of districts in the far west*. San Francisco: Far West Laboratory for Educational Research and Development.
- Schlechty, P. C., & Whitford, B. L. (1988). Shared problems and shared vision: Organic collaboration. In K. A. Sirotnik & J. I. Goodlad (Eds.), *School-university partnerships in action: Concepts, cases and concerns* (pp. 191–204). New York: Teachers College Press.
- Scott, W. R., & Davis, G. F. (2007). *Organizations and organization: Rational, natural, and open systems perspectives*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Sinclair, R. L., & Harrison, A. E. (1988). A partnership for increasing student learning: The Massachusetts Coalition for School Improvement. In K. A. Sirotnik & J. I. Goodlad (Eds.), *School-university partnerships in action: Concepts, cases and concerns* (pp. 87–105). New York: Teachers College Press.
- Snow, D. A., & Benford, R. D. (1992). Master frames and cycles of protest. In A. Morris & C. Mueller (Eds.), *Frontiers in social movement theory* (pp. 135–155). New Haven, CT: Yale University Press.
- Snow, D. A., Rochford, E. B., Jr., Worden, S. K., & Benford, R. D. (1986). Frame alignment processes, micromobilization, and movement participation. *American Sociology Review*, 51(4), 464–481.
- Spillane, J. P. (1998). State policy and the non-monolithic nature of the local school district: Organizational and professional considerations. *American Educational Research Journal*, 35(1), 33–63.
- Spillane, J. P., & Thompson, C. L. (1997). Reconstructing conceptions of local capacity: The local education agency's capacity for ambitious instructional reform. *Educational Evaluation and Policy Analysis*, 19(2), 185–203.
- Spradley, J. (1979). *The ethnographic interview*. New York: Holt, Rinehart, and Winston.
- Stone, D. A. (1988). *Policy paradox and political reason*. Glenview, IL: Scott Foresman.
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.
- Turner, E. O. (2008). *Race talk in the school district: Racial framing and the negotiation of math and literacy policy*. Unpublished manuscript.
- Vozzo, C., & Bober, B. (2001). A school-university partnership: A commitment to collaboration and professional renewal. In T. Ravid & M. G. Handler (Eds.), *The many faces of school-university*



- collaboration: Characteristics of successful partnerships* (pp. 223–233). Englewood, CO: Teachers Ideas Press.
- Wheelan, S. A. (1994). *Group processes: A developmental perspective*. Boston: Allyn & Bacon.
- Williams, D. D. (1988). The Brigham Young University–Public School Partnership. In K. A. Sirotnik & J. I. Goodlad (Eds.), *School–university partnerships in action: Concepts, cases and concerns* (pp. 124–147). New York: Teachers College Press.
- Williams, R. H., & Kubal, T. J. (1999). Movement frames and the cultural environment: Resonance, failure, and the boundaries of the legitimate. *Research in Social Movements, Conflicts and Change*, 21, 225–248.
- Zucker, L. (1991). Postscript: Microfoundations of institutional thought. In W. W. Powell & P. J. DiMaggio (Eds.), *The new institutionalism in organizational analysis* (pp. 103–106). Chicago: University of Chicago Press.

# Collaborative Assistance in a Highly Prescribed School Reform Model: The Case of Success for All

Vicki Park

*University of Southern California*

Amanda Datnow

*University of California, San Diego*

This article focuses on how the Success for All Foundation (SFAF)—the nonprofit intermediary organization that promotes Success for All—works with educators in schools to increase capacity for learning and instruction. Success for All is a comprehensive school reform model that primarily centers on early literacy intervention. Building on research on intermediary organizations and situated learning, we examine how SFAF structures professional development and the types of relationships the organization cultivates with practitioners. At a glance, although much of the theory, strategy, and tools driving the SFAF's approach to school reform seem technically oriented and highly prescribed, our investigation indicates that the deeper process of creating knowledge for school improvement is a collaborative, situated endeavor. Moreover, the study reveals that the process of learning and professional development within the program is a result of the ongoing, dynamic interplay among the SFAF, local conditions, and the broader policy context. Implications for policy and practice are discussed.

Over the past decade, the definition of effective professional development for educators has evolved into one that emphasizes situated learning rather than discrete training sessions removed from the day-to-day needs of the classroom (Borko, 2004; Wilson & Berne, 1999). Much of the literature on situated learning focuses on context-specific learning opportunities whereby teachers within a school relate to one another as peer coaches and as learners in a community

(Cochran-Smith & Lytle, 1999; McLaughlin & Talbert, 2001). Consequently, relationships in such communities are often defined in terms of collaboration, reciprocity, and mutual engagement in developing reflective instructional practices.

Although these types of learning opportunities are critical for teacher professional development, external service providers (i.e., intermediary organizations) have also entered the stage as important actors in building capacity for professional learning and development. This article focuses on how the Success for All Foundation (SFAF)—the nonprofit intermediary organization that promotes Success for All (SFA)—works with educators in schools to increase capacity for learning and instruction. Success for All is a comprehensive school reform model that primarily centers on early literacy intervention. Building on research on intermediary organizations and situated learning, we examine how SFAF structures professional development and the types of relationships the organization cultivates with practitioners. We also analyze how SFAF's work with educators in schools shapes—and is shaped by—policy and market contexts.

## INTERMEDIARY ORGANIZATIONS AND PROFESSIONAL DEVELOPMENT

In developing our framework for this article, we draw on research from professional development as well as investigations on intermediary organizations. Undoubtedly, an integral aspect of increasing capacity for learning and school improvement is professional development. The onslaught of “drive-by” training sessions (Elmore, 2002) that do little to address the specific needs of schools and teachers cannot support the ongoing learning that is required for capacity building (Darling-Hammond & McLaughlin, 1995; Little, 1999). Instead, effective professional development provides teachers with continuous and intensive opportunities to share, discuss, and apply what they are learning with other practitioners (Garet, Porter, Desimone, Birman, & Yoon, 2001; Wilson & Berne, 1999). For this to occur, system-level support needs to be in place. In addition to consistent structured time for collaboration and professional learning, schools need strategies for planning, sharing, and evaluating their efforts. Given the coordination of human, social, material, and technical resources that are required for this scale of change (Hatch, 1998), external service providers are assisting in such capacity-building efforts.

Intermediary organizations recently have emerged as important units of analysis for research on school reform and change (Coburn, 2005; Honig, 2004; McLaughlin, 2006). As noted by McLaughlin, “intermediaries comprise a ‘strategic middle,’ operating between the top and bottom of the implementing system” (p. 220). In particular, McLaughlin highlighted the role of intermediary organizations as bridges between policy and practice. These positions provide intermediaries

access to a diverse array of knowledge and tools, which can help schools or organizations they work with to develop new capacities or to efficiently utilize existing capacities. Hence, intermediaries have the potential to act as knowledge brokers between various levels of the educational system.

The space that intermediaries work within is argued to have several other advantages. First, because they are not government or public entities, intermediaries are often less slow and bureaucratic in their functioning. Second, their role as external organizations allows them the possibility of a singular focus (e.g., instructional reform) and thus increased possibility for innovation in that arena. Finally, intermediaries—and the models they bring with them—can function as sources of stability in schools and districts where there is frequent turnover of leadership or staff (Corcoran, 2003; Marsh et al., 2005).

Intermediaries are a diverse group of organizations. Honig (2004) noted that intermediaries can be categorized along five dimensions. She noted that intermediaries vary along the groups they mediate between, the composition of the organization itself, their location, the scope of their work, and their sources of funding (p. 68). They also often vary on their theories of change (Marsh et al., 2005). These differences notwithstanding, Honig provided a useful definition in delineating the general functions and roles of intermediaries:

Intermediaries are organizations that occupy the space in between at least two other parties and primarily function to mediate or to manage change in both of those parties. These organizations operate independently of these two parties and provide distinct value beyond what the parties alone would be able to develop or amass themselves. These organizations also depend on those parties to perform their essential functions. (p. 83)

Based on Honig's (2004) definition, intermediaries are both independent and dependent of the groups that they interface with and the contextual conditions in which they find themselves. They do not merely mediate or manage change but also create change by virtue of their position as intermediaries and their responses to reform (Coburn, 2005; Cohen & Hill, 2000). Their relationships with different groups allude to a process of negotiation because the parties involved do not necessarily come together with shared norms of communication, behavior, or theories of action.

At the same time, a recent RAND study of the Institute for Learning has found that the capacity of the intermediary and its alignment with local needs can greatly affect partnership success (Marsh et al., 2005). As Marsh et al. (2005) noted, "Without a match between capacity and needs, intermediary organizations run the risk of being relegated to vendor status and seen as tangential to the district's core reform efforts" (Marsh et al., 2005, p. 132). To be successful, intermediaries need to be seen as partners, not vendors.



Moreover, Marsh et al.'s (2005) study noted the importance of intermediaries developing tools that are considered relevant and legitimate in the local context. Context is crucial in determining how intermediaries interact with existing school systems at multiple levels to assist in the development of knowledge, resources, structures, and capacities. No matter how prescribed the reform model, intermediaries are not merely conveyors of information or coordinators of resources but are participants in the co-construction process of reform. Speaking about the complexity of the process, Datnow (2006) stated:

The theory of co-construction rests on the premise of multi-directionality: that multiple levels of educational systems may constrain or enable implementation and that implementation may affect those broader levels. In this view, political and cultural influences do not simply constrain reform in a top-down fashion. Rather, the causal arrow of change travels in multiple directions among active participants in all domains of the system and over time.

In other words, decisions and actions taken at multiple contexts are interrelated and influence one another. Thus, to assess how an intermediary contributes to educational change, its interactions with multiple levels of the system need to be explored.

Consistent with the aforementioned notions, researchers have begun to recognize that learning is both an individual and social activity. Situative theorists build on this knowledge and define learning as changes that occur not only when individuals construct new knowledge but also when people participate in social activities (Borko, 2004; Glazer & Hannafin, 2006; Lave & Wenger, 1995). In other words, knowledge may be developed both individually and collectively, within and across contextual boundaries.

The situative perspective leads us to important implications for research and understanding the process of learning, especially within the context of schools. As Borko (2004) argued, "To understand teacher learning, we must study it within multiple contexts, taking into account both the individual teacher-learners and the social systems in which they are participants" (p. 4). To do so, we need to examine four key elements that make up the learning system: (a) the professional development program, (b) the learners, (c) the facilitators, and (d) the context in which the learning takes place (Borko, 2004).

In this article, we examine the professional development program that is structured around SFA. The learners ostensibly are the teachers and administrators who work with the SFA program. However, given the collaborative and reciprocal nature of the relationship between the SFAF staff and the school, the program developers and trainers are also included as learners in the system. The facilitators are the SFA trainers who work closely with the schools to implement the program as well as the program facilitators who are part of the school faculty. Finally, in

considering the context in which the professional development takes place we include not only the local school system but also the larger policy landscape that shapes the learning system.

## METHODOLOGY

The primary method of this study is case study analysis, with the SFAF itself as “the case.” The case study of SFA was nested within a set of 10 case studies on initiatives that reconceptualize and reorganize the role of research vis-à-vis practice. The broader study was led by Mary Kay Stein at the University of Pittsburgh and Cynthia Coburn at UC Berkeley. Case study methods enabled us to examine the intermediary in a real-life context and allow us to present the perspectives of those actually implementing or working with the program (Yin, 2003). Our research methods were qualitative and involved multiple sources of data including interviews, focus groups, observations, and a review of relevant documents.

Because we were interested in how SFAF works as an intermediary organization and its impact on practice (for a more comprehensive report of our findings, see Datnow & Park, 2006), we studied the SFAF and two SFA schools. In keeping with the tenets of case study research, the sites and participants were chosen purposefully to address our research questions. From SFAF, we interviewed Robert Slavin and Nancy Madden, co-founders of SFA and researchers at Johns Hopkins University. We also interviewed several other SFAF staff including an SFA trainer, the director of training for SFAF, two area managers, an implementation officer, and an individual in charge of policy. Each of these interviews lasted one hour or more. We also observed an SFA Experienced Sites Conference in January 2006.

We also gathered qualitative data during site visits to two SFA schools. The schools were recommended to us by an SFA area manager on the basis that they had been implementing SFA for several years and were in the state of California (important for practical reasons). School A has been implementing SFA for 5 years. School B has been implementing SFA for 6 years and recently became a charter school. Both of the schools are Title I schools, serving large numbers of low-income students. The majority of the students in both schools are Hispanic. Both are large schools serving more than 1,000 students each. The schools are located in different school districts—one in a very large urban district and one in a midsize district.

Our data collection at each school involved interviews, focus groups, and classroom observations. At both schools, we conducted interviews with the school principal. At School A, we also interviewed the SFA facilitator. At School B, we interviewed the assistant principal who oversees SFA and several lead teachers who serve as SFA facilitators. Because the school is so large, they had individuals

at each grade level helping to facilitate SFA rather than just one facilitator. At School A, we conducted a focus group with four regular classroom teachers from different grade levels. At School B, we also interviewed three teachers from various grade levels. All interviews were tape-recorded and transcribed verbatim. In both schools, we observed classroom instruction in five classrooms per school during the SFA 90-min reading period. We also took field notes during classroom observations. Although this school-level data were important for the larger study, we rely on it only to a limited degree in this article, as the SFAF-level interviews were more important for this analysis.

Our qualitative data analysis activities have included coding the interview transcripts based on the research questions guiding this study. Interviews were coded with the aid of qualitative data analysis software called Hyper Research. Our review of documents included reports that details SFA's history (Slavin & Madden, 1998; Slavin, Madden, & Datnow, 2005), articles written about federal policies (Brownstein & Hicks, 2005; California Reading First Technical Assistance Center, 2005; Manzo, 2005b), and newspaper articles about SFA. For the purposes of confidentiality, pseudonyms are used for all school, person, and place names. However, we are unable to keep the identities of Robert Slavin and Nancy Madden confidentially, given their role as leaders and founders of the SFAF.

## FINDINGS

Our broad purpose in this article is to explain how the SFAF builds the capacity of educators in SFA schools. Our findings are divided into several sections. First, we provide an overview of SFAF's approach to professional development. Second, we explain how SFAF trainers work to develop knowledge in schools, explaining the relationship between SFAF and educators in schools. Finally, we discuss how the broader policy context has shaped how SFA works with schools.

Before we discuss our findings, we provide a brief background on the SFA model and the structure of the SFAF itself. As previously noted, SFA is a comprehensive school reform model that primarily centers on early literacy intervention. Originally developed by Robert Slavin, Nancy Madden, and a team at Johns Hopkins University, the program is currently based at the SFAF in Baltimore. Over the past 15 years, the number of schools implementing SFA has grown substantially, with the SFAF working with approximately 1,500 schools in 46 states, as well as assisting in related projects in five other countries. Most SFA schools receive Title I funds and serve large numbers of low-income and minority students. Although the majority of SFA schools are at the elementary level, the SFAF also works with a smaller number of preschools and middle schools.

The SFA program is an interesting case to study because of its high-profile success in scaling up to more than 1,500 schools and its status as one of the



more prescriptive comprehensive school reform models. There is a great deal of "formal" research knowledge on the effects of SFA and evidence that this research knowledge informs the program and its continual development (Slavin et al., 2005). There is also an inherent assumption in the program, by virtue of its specification, that knowledge for school improvement can in fact be created by external groups and transported across contexts.

The program's instructional approach to increasing student achievement revolves around three main strategies: research-based instructional techniques for teaching reading, cooperative learning, and the use of data and ongoing assessment. Major components of SFA include: a 90-min reading period every day; the regrouping of students into smaller, homogeneous groups for reading instruction; quarterly assessments; and one-to-one tutoring. The SFA reading curriculum is composed of an Early Learning program for pre-kindergarten and kindergarten students; Reading Roots, a beginning reading program; and Reading Wings, its upper-elementary counterpart. There are both English and Spanish versions of the program. In addition, the SFAF has developed writing, math, and social studies programs.

The program takes an aggressive approach to changing teaching and learning. As a result, SFA is highly specified and comprehensive with respect to implementation guidelines and materials for students and teachers. Almost all materials for students are provided, including reading booklets for the primary grades, materials to accompany various textbook series and novels for the upper grades, and activity sheets and assessments for all grade levels. Teachers are expected to follow SFA lesson plans closely, which involve an active pacing of activities during the 90-min reading period (Madden, Livingston, & Cummings, 1998).

The SFA model also takes a specified approach to the adoption process. The SFAF requires that the majority of a school's teaching staff to participate in voting for program adoption before they provide them with materials and technical assistance. The program also asks that schools employ a full-time SFA facilitator, organize a Solutions Team to help support families, and organize biweekly meetings among Roots and Wings teachers. The principal of a SFA school is responsible for ensuring staff motivation and commitment to the program as well as adequate resources to support it. The role of the SFA facilitator is to ensure the quality of the day-to-day implementation of the program by supporting teachers, monitoring the progress of all students, and managing assessments and regrouping efficiently (Madden et al., 1998). Implementation of the program is supported through ongoing professional development from SFA trainers and through local and national networks of SFA schools (Slavin & Madden, 1998).

The SFAF houses teams of researchers, area managers, trainers, a production section, and a business section. Robert Slavin acts as the chairman, and Nancy Madden is the president of the foundation. Both are extensively involved with program development and directing of SFAF. One staff member describes the



organization as “vaguely hierarchical,” and many of the people we interviewed from SFAF had difficulty remembering their titles because they were taking on multiple tasks and roles. During the period of our study, the number of staff totaled approximately 220 people. SFAF employs regionally based trainers, some of whom work in offices and others who work out of their homes. The Foundation recruits trainers from schools, usually former SFA teachers and facilitators, who have expert knowledge in how the model works in a particular local context. Overseeing the 150 trainers are 15 area managers, who also deal with district relations and respond to trainers questions regarding school adaptations of SFA. Area managers are supported by a team of expansion and outreach staff. Two “implementation officers” oversee the area managers and outreach personnel. Although some staff members are located at the foundation offices in Baltimore, others are spread around the country.

### *SFAF's Approach to Professional Development*

In this section, we detail the ways in which SFAF structures its professional development for practitioners and the pivotal role that SFAF trainers play in transferring and sharing knowledge. We then conclude with an analysis of the ways in which SFAF and its relationship with schools has evolved as a result of the dynamic interplay among multiple contexts.

SFAF's approach to professional development adheres to a coaching model built upon the work of Showers and Joyce (1996). In this model, there are four main components: the developing knowledge of instruction and curriculum, understanding the underlying theory of skills and tools utilized, modeling of skills, and the practice of skills (Joyce & Showers, 2003). Although the Foundation has incorporated all these elements into their curriculum-based professional development program, different components have been stressed to a varying degree. Overall, SFAF's approach toward their professional development program emphasizes situated learning and sharing practices across contexts.

As part of their approach, the SFAF makes structured professional development a cornerstone for building knowledge of teaching and learning. Extensive professional development includes initial training in reading strategies, instructional delivery methods, and monitoring of student progress using assessments. Each school receives telephone consultations two or three weeks after trainings to answer questions regarding implementation. There are two on-site support visits over the year to observe students' strategy use in classrooms, to meet with teachers and administrators, to review data on student progress, and to set new goals. In addition, there are six follow-up telephone meetings to provide teachers with further training and support for their implementation of the program. The meetings are held on a quarterly basis to answer teachers' questions and help with

troubleshooting, goal setting, and assessment issues. In general the foundation has unlimited, informal telephone support for all staff members. Beyond professional development associated with initial training, schools also can participate in conferences held by the foundation, which are divided into Experienced Sites and New Sites.

With their emphasis on continuous improvement, the SFAF asks schools to commit a significant portion of their human and financial resources to professional development activities. Madden described the endeavor as “interaction heavy” because, SFAF staff typically spend approximately 26 days at the school site in the first year of implementation. Undoubtedly, SFAF has more interaction with some schools than others, depending on their maturity in program implementation and on the number of days that schools contract with SFAF for training. Continuing implementation sites can contract for as few as two days of training (the minimum) to as many as they need. SFA trainers across the country work with anywhere from 5 to 20 schools each, depending on the number of contracted days per school. “Every school is assigned a point trainer and sometimes even a co-point, and that’s their first line of communication,” explained a SFAF staff member. The two schools in this study were mature implementation sites but continued to invest considerable amounts of time and resources in ongoing training. Thus, the close contact SFAF has developed with schools through professional development and a goal-focused approach, means that the strategies used to help schools increase student achievement have become a more dynamic, localized and contextualized process.

School leadership development is also a noted feature of the SFAF’s coaching model. At both schools in this study, we found that the success of SFA relies on the principal working as a program advocate, whereas the facilitator acts as the program manager and provider of staff support. This is confirmed by prior research on SFA (Datnow & Castellano, 2000). Decisions regarding the allocation and coordination of resources, especially the number of days to contract for training, and the degree of on-site professional development, are made at the school level, and thus the role of principal is a key factor in their determination. The fidelity of implementation is led by school administration, with some principals using a more flexible approach than others (Datnow & Castellano, 2000). The Leadership Academy provided by, SFAF provides opportunities for school leaders to share and reflect on practices as well as to set and monitor their progress within and across districts. Typically, schools send three or four representatives from each school to a training session. The first year involves eight sessions, followed by four sessions in the second and third year. Each session is structured around content aimed at leadership development and goal-focused planning.

The SFAF has always focused on building instructional knowledge, modeling, and practicing in their work with teachers. Professional development is composed of three main components: intensive initial training, ongoing coaching, and

goal-focused planning for all program members (Success for All Foundation, 2005). In terms of curriculum training, the SFAF has shifted to becoming much more detailed. For instance, cooperative learning has always been a centerpiece of the program, but specified routines and student procedures have been developed to enable teachers to provide more explicit modeling. The director of the SFA Training Institute believed that people generally understand cooperative learning standards but do not necessarily know what it looks like in practice. The emphasis on explicit modeling and learning for teachers seems to be an extension of learning theories applied to students. For example, cooperative learning is not only utilized in student learning but also applied in teacher development. Madden shared that “having teachers work together on these instructional concepts makes it easier for them to grow and refine their skills.” However, cooperative learning does not automatically transform in to teacher-driven inquiry; emphasis is placed on using the same instructional processes, training, and similar language as the foundation for teaching. During teacher learning communities, teachers meet monthly to talk about their instructional practices. Ideally, they would have a corresponding video segment (e.g., phonics) that could act as a conversation starter for teachers to discuss and relate back to their own practices. These types of reflections then lead to plans for implementation, evaluations, or modification. Through speaker phones, one of the trainers might mediate the meeting to help build more in-depth understanding of particular skills that teachers are introducing to their students.

The use of video underscores the importance of tool development in aiding teacher learning and practice. These codified tools provide the basis for shared knowledge and a common language, as well as translating abstract concepts to concrete practices. As mentioned earlier, the importance of developing instructional materials revolves around the premise that teachers need tools that are accessible and easy to use. Detailed teachers’ manuals, video-taped modeling strategies for students and teachers, computer programs (e.g., *Alphie’s Lagoon*, which is a computer-based intervention program), and data analysis software are important foundations for training and communication.

In addition to the “model, teach, and practice” approach, SFAF increasingly emphasizes an understanding of theory behind the tools. In evaluating the quality of the program’s implementation, they found that schools and trainers were overly focused on visible details of the implementation rather than the theory underlying their use. Superficial engagement with the program seemed to result in compliance to mandates and fidelity to implementation but did not necessarily translate into enhanced student outcomes. The new theory is that the effective use of tools is driven by understanding the purpose and conceptual development behind the tool. Madden shared,

At first we spent much more time on the activities and just getting teachers fluent with the activities so that they would be utilizing the concepts and lately we’ve sort



of presented it more conceptually first, and then gotten down to “how to,” and I think that’s where we get a much richer implementation by the teachers if we can get them to do both of those things.

One school-based facilitator observed,

SFA has changed, too, along with us. They’ve loosened up a lot. When they first started, it was very “you follow this routine, you get it done in this amount of time.” And with their research they have found that it doesn’t really matter if your knees are touching and you’re facing each other in partner reading.

Thus, SFAF found that it was important for teachers to understand the theory behind the tool in order for it to be utilized more effectively or adapted to fit the needs of students. For mature school sites, such as the ones in our study, the notion of fidelity of implementation was now balanced by a “goal-focused” implementation. That is, rather than focusing on adhering to routines, procedures, or instructional strategies rigidly, SFAF encourages practitioners to think more reflectively about their practices and to utilize whatever is most effective at improving student achievement. Consequently, they have started to emphasize this in their professional development approach.

### *The Role of Trainers in Knowledge Development*

Understanding how SFAF is structured with respect to the training and support of schools is important to making sense of the process of how knowledge development works within the organization. The role of SFA trainers in the development of the knowledge of the model is pivotal because the network of trainers acts as an important mechanism by which best practices are shared and disseminated. Before describing the networking of trainers and the process by which knowledge is disseminated, a brief description of trainers’ backgrounds further illuminates how SFAF draws upon trainers’ working knowledge of learning and teaching.

In the early development of the program, the researchers themselves worked as trainers. As the program grew, the SFA team quickly began to see the value of recruiting trainers who had extensive background working as practitioners. Trainers are minimally required to have a bachelor’s degree, professional certification in their designated field and five years of experience in education. Individuals with experience in using the SFA program are preferred. New trainers participate in a three-week “train-the-trainers” program organized by the New Trainer Institute. In addition, new trainers also co-train with more experienced trainers. As noted by the developers, “The support of an experienced trainer is modeled on the coach-teacher relationship that is integral to the successful implementation of the curriculum” (Success for All Foundation, 2005). Besides having more years of



experience in education, senior trainers also have two to four years of SFAF training experience. Staff interviewed for this report have a wealth of clinical experience working in schools and/or related fields. Regional area managers have knowledge of local contexts because they have experience working as teachers, facilitators, or support providers. Staff members located at the foundation have a broad range of knowledge as well. For example, one of two vice presidents for field operations was a former mental health therapist working with students before becoming a trainer for SFA Family Support. Additionally, the Foundation's Education Policy and Constituent Relations Manager had previous experience working as a reporter and researcher before coming to SFAF.

Networking between trainers through their own professional development from conferences and meetings acts as one of the main sources of knowledge transfer and dissemination between schools. Annual conferences are held for trainers where they convene to share strategies for working with schools and discuss successful and unsuccessful program adaptations they have observed. One trainer mentioned that in a given year she will participate in a week long Experienced Trainer Institute, multiple (regional) team meetings, and national training institute sessions. During these conferences, trainers also engage in discussion about current research that is related to SFA, but not on SFA per se. For example, we heard that foundation staff might refer to new research on strategies for teaching English language learners. Trainers then use this research when they are meeting with school educators to help them understand why particular program components are necessary. An area manager noted that reflection is an integral piece of program success. Trainers are expected to conduct a case study on a school, focusing on success and challenges with meeting goals. Then they meet annually in Baltimore to review each case study and what was learned. One trainer explained, "Our consulting services are really about coaching reflective behavior and reflective thinking" so that teachers can articulate and focus on strategies that improve student learning.

When components are developed or there is an added adjustment, the SFAF sends updates through its monthly newsletter referred to as *Trainer Times*. Furthermore, the Training Institute acts as a depository/inventory of questions and feedback from trainers. Informally, trainers send out e-mails to the national network for sharing and troubleshooting. One area manager explained the process:

We internally share what's working. . . . If we're working with a school that has a certain area of focus, and we need some ideas, then there's a network with a lot of us out there that we can say, "hey, has anybody ever worked with this, do you have anything that will support this school?" They send an e-mail, they you get tons of responses and ideas, so the networking is really that powerful.

Informal networking and sharing between schools produced changes at the program level. Although much of the development of tools is structured and

systematic, others arose serendipitously, such as the creation of an online data base analysis used by member schools. One of the facilitator's husband, who had knowledge of computer programming, developed a spreadsheet program that slowly spread in use throughout different schools sites by word of mouth before it was refined and institutionalized within the foundation. The SFAF now has a sophisticated data analysis program that is accessible to schools through the foundation's Web site. This technology permits schools to view student reports and enables them to sort the data based on a variety of filters (e.g., subskills, Adequate Yearly Progress [AYP] subgroups, homeroom teacher, and reading group).

Regarding the process of communication between schools and foundation staff, a SFAF area manager explained,

It usually goes point trainer, me, and then whoever needs to be contacted within the Foundation. . . . We have several different people for each component or skills area—technology, ELD, whatever it is. . . . And that's one thing we let teachers know, you know, schools know, we have this huge foundation behind us, you know, if I can't get you the answer, there's somebody back there that designed this that can tell you exactly why it is the way it is.

As this statement implies, the foundation was seen by this trainer as a storehouse of knowledge that could be accessed as needed. Along these lines, there were additional, relatively new efforts at gathering knowledge from educators about their experiences with SFA. First, SFA trainers were available on speakerphone to address questions and gather feedback during teachers' monthly "teacher learning community" meetings that occurred in each school. Teachers also engaged in discussion about their practice at biweekly meetings at each school site.

Face-to-face interactions between schools and trainers provide a mechanism by which schools were able to address needs specific to their local context. Trainer-school relationships were central bridges between the program developed by SFAF and the actual implementation. Trainers made abstract concepts such as reading strategies more concrete and provided models and examples. Overall, the ongoing training was seen by educators as an integral part of building their own knowledge for school improvement. One SFA facilitator we interviewed explained that schools received support based on their self-identified areas of need, rather than generic support: "When [our trainer] comes in to sit with [the principal], she asks us, 'What are the needs of your school? What training do you want from us? What support?' So it's not just them dictating what support it's going to be." Another statement by a principal gives insight into the level of collaboration between SFAF trainers and educators in her school: "I think we have always felt a give-and-take and that we are accepted as peers and colleagues . . . that they are interested in what we say and that there's a response to that."

The types of relationships between the schools and trainers were also indicative of the depth of engagement with program implementation. A positive, trusting relationship facilitated a deeper level of engagement with the program components as well as allowing for continuous feedback between the two parties. Madden described trainers as “really expert teachers . . . who just had real strong concepts about teaching and could communicate that to other teachers, have the respect of other teachers and could really work with all of them to be really good coaches.” Madden’s comments indicate that trainers not only have acquired deep knowledge about curriculum and instructional delivery but also have interpersonal knowledge that enables them to communicate effectively with various participants. This appreciation for interpersonal knowledge was expressed by both schools participating in our study. Educators acknowledged how important it was for the trainers to “match” the needs of their school teachers and students. As one administrator explained, “They’re like a perfect match for our school. The people, their personalities, their backgrounds in bilingual [education] you know, for ELD [English Language Development]. . . . I think they tried to match the person to the school because they are just perfect for us.” Therefore, teachers and principals felt that the trainers were well matched to their schools in terms of expertise and background.

As previously noted, the SFAF strives for a collaborative relationship with its schools but continues to focus on the importance of providing educators with detailed instructional delivery guidelines for program implementation. Slavin remarked, “We’re trying to get a proper balance [between implementation fidelity and adaptation], but I think if you’re truly serious about change, about having teachers use research based practices every day, you’ve got to be pretty explicit and pretty well thought out to have that take place.” He added, “Part of our theory of action has to do with trying to get away from the script but still have teachers understand what they are doing by showing things directly to kids and then model that strategy for the teachers.”

Frequent dialogue between school educators, trainers, other SFAF staff, and the SFA directors informs the development of the program. This was a change from the early development of the program when far fewer people, in a far less diverse set of roles (e.g., mainly researchers), were informing development. The educators we spoke with believed that they had an open dialogue with SFAF staff. As one facilitator said, “If we have a question . . . they are really good about emailing us back an answer or calling us back with an answer.” The topics of these discussions might range from implementation issues, successes or problems with particular program components and the degree to which SFA was helping meet schools’ goals. The process is such that trainers gather information about how the program is working when they visit local schools, and they share this knowledge with other foundation staff on a regular basis. At one school we visited, the principal, teachers, and SFA facilitator all mentioned taking part regularly in this type of dialogue.

A teacher observed that some program changes resulted from their feedback and commented:

It's kind of like formal versus informal research, because some of the component meetings we have with our trainers. We sit down and she asks us questions, "Are you comfortable? Do you need more training? What do you like? What don't you like?" And so she's kind of doing informal research, you know.

As the comment implies, the teachers we interviewed very much felt as though they were an integral part of the continual development of the SFA model. In addition, SFAF staff conduct phone interviews with educators in particular schools take place, particularly when the foundation staff is interested in finding out how a new component of SFA is working.

Madden indicated that some of the "best changes" to the program are a result of the feedback from schools. She described the process:

A sort of situation arises where a school has some feedback to give us, you know, they want to let us know that they've figured something out . . . or they're having a special problem, and then maybe I'll go out and take a look at it and spend some time to really get out and understand it, or [another staff member] will go and work on what is the issue, how can we learn from it, or how can we help with it, and then we take that back to the development organization and say what can we do realistically to take to use this information.

These examples point to the ways in which SFA has substantially broadened its notion of what counts as valid knowledge in the development of the program. The fact that SFA has changed its stance over time underscores how organizations like SFAF are not static entities but rather change and learn as their programs grow and their own knowledge about school improvement deepens.

### *The Interplay of Multiple Contexts and SFA*

It is important to consider the role of the broader policy context in relation to SFA, as it has been instrumental in shaping how the reform model and the capacity building processes within SFAF have changed over the past few years. As we argue, the policy context has both enabled and constrained SFA—and more recently has become part of the "knowledge" of what constitutes SFA. Federal policies such as the Comprehensive School Reform program and the changes in Title I laws helped SFAF expand exponentially in the 1990s, posing challenges in meeting rising demands and implementation problems. The current climate has been more stable with SFAF providing services to roughly the same number of schools for the past several years. SFA schools have had the program for a median



of five years (Slavin et al., 2005). This stability has enabled SFAF to develop depth of experience within its staffers and schools.

Whereas the federal policies of the 1990s advantaged SFA in terms of scale up, recent federal policies have generally been constraining. In particular, policies accompanying No Child Left Behind (NCLB) led to numerous challenges. The push for statewide curriculum alignment and the emphasis on testing have made it difficult for schools to implement a reform model like SFA which emphasizes early intervention, prevention of reading difficulties, and whole school change. One SFAF staff member explained,

The pendulum has swung the other way, when there was a time a few years ago on parent involvement, community involvement, how to intervene effectively with kids, how to address issues school-wide in a preventive manner . . . With NCLB, now prevention is rarely addressed because nobody has any time, because it's all about how kids do on a test on a given day.

Another challenge for the SFAF has been the Reading First Initiative, a federal policy regarding reading instruction established under NCLB, Title I. Reading First is the \$1 billion-a-year federal program that provides funds for scientifically research-based reading programs targeted for disadvantaged children in the primary grades. Despite NCLB's more than 110 references to scientifically based research and its emphasis on evaluations that use random assignment or quasi-experimental designs, local education agencies applying for Reading First grants have typically been unsuccessful when including SFA as part of their proposals. Slavin explained, "We've had a real disaster with Reading First, which everybody in America, everybody in the education world, believes is being enormously beneficial to [SFA] . . . that it's talking about research-based practice. But, in fact, it has been extremely damaging." This is attributed to the policy favoring traditional basal readers. This is, considered ironic, as Reading First purports to support scientifically based practice.<sup>1</sup> Coinciding with the implementation of Reading First, the reading text adopted in the state of California, has constrained SFA, as it allows for the adoption of only one of two reading series (Houghton Mifflin and Open

---

<sup>1</sup> States apply to the Department of Education (ED), which allocates grants based on the number of students living below the poverty line. States, in turn, distribute funds to local educational agencies on a competitive basis. A panel of experts reviews state proposals, which require a plan to implement a scientifically reading-based instruction program and monitor student progress. Once approved, states are required to submit annual implementation and progress reports. The ED's panel of experts and consultants oversee the implementation process. The ED has set up a National Center for Reading First Technical Assistance to aide states and districts with implementation, including assistance in reviewing programs, materials, and assessments. The assistance is divided by three regional centers operated by the University of Oregon College of Education, The University of Texas at Austin, and Florida State University.

Court), which educators have interpreted as excluding SFA as an option (Manzo, 2005b). In fact, SFA can be adapted for use with both of these reading series. However, most educators and policymakers are not aware of this, resulting in situations where schools are told by Reading First officials that the two programs are incompatible.

In 2001, the year before the first distribution of Reading First, 200 schools adopted SFA. Since then, several hundred schools have dropped SFA and only five Reading First schools adopted the program. In recent years, the foundation had to lay off more than 300 staff members (Manzo, 2005a). Madden explained,

Essentially, we have been pushed out of even very successful Success for All schools based on the current policy because “we don’t fit Reading First” and the reason we don’t fit is because research . . . I mean, Reading First has sort of developed this set of, you know, rules that are not based on the legislation and are not even based on the guidelines, but that everyone is being held to, with the threat they’ll lose their funding.

For example, in one state, schools were told they could not use SFA because they were required to use instructional centers, which are not part of the SFA program.

The movement of school districts toward curricular uniformity and coherence has posed additional challenges for both SFAF and SFA schools. As districts have moved toward mandating literacy programs across entire districts, this has sometimes meant the demise of SFA (Datnow, 2005). Several SFA staff mentioned how the shift from site-based decision making in the 1990s to district-based decision making in the early 2000s has influenced their approach to program adoption and support. Because SFAF believes in the importance of staff buy-in for successful program implementation, it typically required the majority of school staff to vote in favor of program adoption before agreeing to work with the school. The two schools participating in our study adopted the program before the push for district coherence. They indicate that there is a constant challenge to justify the continuation of SFA, despite rising student test scores. One of the principals of the school site commented,

It’s frustrating sometimes for those of us who are principals at SFA schools, when we go to the district things [meetings] and they’re going, this is what you should be doing, and we’re going, we are, you know, but you don’t recognize that’s what the program is.

One school in our study, located in a large urban district that has mandated the Open Court reading program, actually converted to a charter school in order to have the freedom to continue with SFA. They use Open Court texts with SFA strategies.

The need to meet local and state accountability demands has restricted the ability of schools to make these types of curricular decisions on their own, leading the SFAF to strategically change both its approach to the marketing of the program and relationships with school districts. Gaining support from districts has been increasingly critical to SFA's sustainability in school sites. SFAF staff members indicate that they now have to gain the support of two different types of audiences to have program adoption. One staff member noted,

We sort of have a district-level awareness presentation, where all they want to hear about is the data, data, data. Then you have teachers that are like, can we get past the data . . . let me see these materials, let me see what I'm going to be using with my kids.

District wholesale adoption of SFA does not necessarily translate as a positive outcome for SFAF. Madden noted,

Over the last seven or eight years, we got involved with more schools who were not showing growth and who were being required to do something, and that puts you in a different relationships with schools. We still ask schools to vote and to take on the program voluntarily, but in some cases the schools feel like they were coerced to vote for Success for All. That changes the chemistry.

Although policy shifts have posed challenges for the funding and development of the program, they have also led to some important, well-received changes within the SFA model and in how SFAF works to build the capacity of educators in schools. As a result of the changes in federal policy, the SFA model itself has changed. Knowledge about how to interpret federal policy has become part of what SFAF offers to schools. This has also changed the way that SFAF works with schools. SFA trainers are now serving as policy mediators or policy knowledge brokers with respect to NCLB guidelines. This appears to have increased the level of collaboration and community between trainers and educators, as educators now see trainers as allies in their quest to meet accountability demands. At the same time, SFAF has also found ways to work more flexibly with schools, particularly with respect to implementing the whole model versus particular components.

SFAF has an organized and institutionalized way of developing policy knowledge. First, the policy changes at the federal and state levels have been so numerous and significant in the past few years that the SFAF has appointed a person in charge of keeping up with federal and state policies. An SFAF staff person we interviewed called this person the "Policy Master," though her official title is Education Policy and Constituent Relations Manager. One of her main responsibilities is to help SFA schools and districts make sense of the AYP accountability mechanism in

NCLB and to help them set goals accordingly. She explained, “In order for us to provide the level of service and quality we need to with our schools, we’ve had to get very state-specific and know the ins and outs of the state interpretation of federal guidelines.” An area manager mentioned that she often knew about policy changes from the Policy Master well before her schools did. During SFAF “area manager camps,” part of the intensive training session includes a 90-min session just on Reading First. Area managers also receive professional development in state standards and any additional updates on state changes, which are then passed on to trainers. One area manager shared, “I’m always researching and conveying to any consultant walking into the school, so that they understand it completely, and can support the school and also get the achievement results.” She herself spent time in meetings and hearings at the state department of education so that she could keep abreast of state policy changes. Trainers also request schools to share or pass along new policy memos, which are then spread throughout the SFA network. Thus, the advent of NCLB has meant that trainers now serve as policy mediators, helping schools gain the knowledge to meet state and federal mandates. An area manager explained the shift in how trainers interface with schools:

Five or six years ago, it was like, “Yeah, it looks good; you’re asking the right kinds of questions, maybe you should try this.” It’s much more global now, and it’s all about aligning. . . . We’re very aligned to No Child Left Behind, we bring them the information, we help them interpret what it really means, “This is what the federal law is saying, let’s look at what California is saying, and see where you fit.” Like the last year and a half, really, it was mostly training full staff and even district folks on what No Child Left Behind really was about because the districts could not keep up with it very well. They just didn’t have the time or the funds or whatever.

Trainers helped schools understand how they can meet NCLB mandates, apparently providing them with knowledge that their districts did not. As one school administrator explained, “If we move those 124 [students in terms of achievement] then we’re going to make Safe Harbor [with NCLB]. Well, we never knew this. None of the principals around here knew this. The district never told them that. And we found this out through SFA.” Trainers also work with educators to help them use SFA to meet state curriculum standards. This process is much more localized, based on individual school needs. As one principal explained, “When [our trainer] has come out, she has met with grade levels and we have done an item analysis by question, by standard . . . okay, what parts of SFA can address this standard and where.” SFAF has also recently developed benchmark SFA assessments that schools can use five times a year. These benchmark assessments, called 4Sight, are linked to state assessments. Although obviously reflecting a change in response to the policy climate, this



change was also made very much in response to the requests of educators, who wanted assessments that related better to what they were being measured on by their states. Slavin explained, “That was based very much on the comments from practitioners.”

In addition to changes in the type of relationship they have with schools, the foundation has also developed another strategy with regards to marketing of the program. SFAF stresses that they are a comprehensive school reform model rather than a publishing or textbook company; yet they are continually invited by states and districts to compete with textbook publishers. Because schools that adopt the whole SFA model need to invest a great deal of resources and time in restructuring their schools and this process can be hampered by district constraints on school-site decision making, SFAF has “unbundled” its program components. In response to the numerous changes in the policy climate, SFAF now allows schools to purchase some components, such as the Early Learning program, a la carte rather than adopting the whole comprehensive school reform model now referred to internally as “Classic SFA.” Still, a large part of SFAF’s theory of action regarding successful implementation of the program emphasizes the “coaching model” and thus require schools to purchase professional development alongside materials. As one staff member shared,

We still adhere to the coaching model, so a school that just may want to buy *Fast-track Phonics*, would not be able to purchase our materials, it still requires ongoing, on-site support by us, because the research . . . there is just very clear evidence of effectiveness there, that just selling people materials is not effective, unless you’re a really high-performing school.

Similarly, funding constraints have led SFAF to think more creatively about communicating with schools. For example, instead of contracting 15 days of professional development, the SFAF might allow for eight days with extra follow-ups through phone and e-mail. As this discussion suggests, SFAF’s approach to building capacity in schools has both been shaped by the policy and market contexts in which they work. They have attempted to incorporate policy knowledge into their professional development services in an effort to better respond to the school improvement market of today. The role of the SFA trainer now includes bringing knowledge about federal and state policies to SFA schools. Now trainers work in a coaching capacity with local educators to help them use SFA most productively to meet state and federal policy demands. As such, not only do trainers work in building educators capacity for teaching reading but they also work as consultants in overall school improvement planning.

## CONCLUSION

At a glance, although much of the theory, strategy, and tools driving the SFAF's approach to school reform seem technically oriented and highly prescribed, our investigation indicates that the deeper process of creating knowledge for school improvement is a collaborative, situated endeavor. Undoubtedly, the SFAF makes structured professional development a cornerstone for building knowledge of teaching and learning. The SFAF has always focused on building instructional knowledge, modeling, and practicing in their work with teachers. Over time, however, the organization has found it increasingly valuable to combine explicit, detailed modeling for teachers with an emphasis on an understanding of the theory behind the tools. SFAF focuses less on measuring fidelity of implementation and more on helping educators to think more reflectively about their practices and to utilize tools that are most effective at improving student achievement.

SFAF's approach toward professional development program also emphasizes sharing practices across contexts. The knowledge of SFA trainers, many of whom were former SFA teachers, is also integral to the continual development of the model and its capacity building strategies. The trainers help bring knowledge of teaching and research together for educators in schools. They also serve to bring information from schools back to SFAF. A positive, trusting relationship between educators in SFA schools and trainers facilitates stronger engagement with SFA as well as allowing for continuous feedback between the two parties. As such, frequent dialogue between school educators, trainers, other SFAF staff, and the SFA directors informs the development of the program.

Moreover, our research reveals that the process of learning and professional development within the program is a result of the ongoing interplay among the SFAF, local conditions, and the broader policy context. It consists of collaboration, negotiation, and conflict along several dimensions—relationships between schools and the SFAF, local and state contexts, and federal educational policy. The interconnections among these dimensions shapes SFAF's strategies toward knowledge development. In turn, SFAF influences the educational policy landscape and the school reform process, including the larger debates surrounding the contested definition of scientific knowledge and the usability of educational research.

Although this study provided some insights into how intermediaries assist in the learning of educators, further research is needed in this area. It would be particularly fruitful to further examine how intermediary organizations make decisions about professional development, much in the way we have attempted to do so here. That is, as intermediary organizations initially plan and subsequently reshape their professional development components, does research on "best practices" in professional development come into play? How do their experiences working with schools—the trial and error of their daily work—factor in? And how do market and policy contexts, particularly accountability pressures and shrinking budgets, shape

their work? Continuing to gather more data on these questions would help build a greater knowledge base for researchers and help the intermediaries themselves as they seek to improve their work with schools.

## REFERENCES

- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researchers*, 33(8), 3–15.
- Brownstein, A. & Hicks, T. (2005). Special report: Reading First under fire. *Title I Online*. Retrieved October 5, 2005 from [http://www.titleionline.com/libraries/titleionlin/news\\_desk/tio050826.html](http://www.titleionline.com/libraries/titleionlin/news_desk/tio050826.html)
- California Reading First Technical Assistance Center. (2005). *California Reading First initiative*. Sacramento, CA: Reading Lions Center. Retrieved October 18, 2005, from <http://www.calread.net/>
- Coburn, C. E. (2005). The role of nonsystem actors in the relationship between policy and practice: The case of reading instruction in California. *Educational Evaluation and Policy Analysis*, 27(1), 23–52.
- Cochran-Smith, M., & Lytle, S. (1999). Relationships of knowledge and practice: Teacher learning in communities. *Review of Research in Education*, 24, 249–305.
- Cohen, D. K., & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record*, 102(2), 294–343.
- Corcoran, T. (2003). *The Merck Institute for Science Education: A successful intermediary for education reform* (CPRE Research Rep. Series RR-054). Philadelphia: Consortium for Policy Research in Education.
- Datnow, A. (2005). The sustainability of comprehensive school reform in changing district and state contexts. *Educational Administration Quarterly*, 41(1), 121–153.
- Datnow, A. (2006). Connections in the policy chain: The case of Comprehensive School Reform. In M. I. Honig (Ed.), *New directions in education policy implementation: Confronting complexity* (pp. 105–123). New York: State University of New York Press.
- Datnow, A., & Castellano, M. (2000). Teachers' responses to Success for All: How beliefs, experiences, and adaptations shape implementation. *American Educational Research Journal*, 37(3), 775–799.
- Datnow, A., & Park, V. (2006). *Scaling up the knowledge of school improvement: The case of Success for All*. Working draft of final report to the MacArthur Foundation.
- Darling-Hammond, L., & McLaughlin, M. (1995). Policies that support professional development in an era of reform. *Phi Delta Kappan*, 76(8), 597–604.
- Elmore, R. F. (2002). *Bridging the gap between standards and achievement: The imperative for professional development in education*. Washington, DC: Albert Shanker Institute.
- Garet, M., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915–945.
- Glazer, E. M., & Hannafin, M. J. (2006). The collaborative apprenticeship model: Situated professional development within school settings. *Teaching and Teacher Education*, 22, 179–193.
- Hatch, T. (1998). The differences in the theory that matter in the practice of school improvement. *American Educational Research Journal*, 35(1), 3–31.
- Honig, M. I. (2004). The new middle management: Intermediary organizations in education policy implementation. *Educational Evaluation and Policy Analysis*, 26(1), 65–87.
- Joyce, B., & Showers, B. (2003). *Student achievement through staff development*. Nottingham, UK: National College for School Leadership. Retrieved September 17, 2007 from <http://www.ncsl.org/media/F7B/94/randd-engaged-joyce.pdf>

- Lave, J., & Wenger, E. (1995). *Situated Learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
- Little, J. W. (1999). *Teachers' professional development in the context of high school reform: Findings from a three-year study of restructuring schools*. Washington, DC: National Partnership for Excellence and Accountability in Teaching.
- Madden, N., Livingston, M., & Cummings, N. (1998). *Success for All, roots and wings principal's and facilitator's manual*. Baltimore: Johns Hopkins University.
- Manzo, K. M. (2005a, June 22). Complaint filed against reading initiative: Success for All officials request investigation of federal program. *Education Week*, 24(41), 3.
- Manzo, K. M. (2005b, June 8). States report Reading First yielding gains. *Education Week*, 24(39), 1, 17.
- Marsh, J. A., Kerr, K. A., Ikemoto, G. A., Darilek, H., Suttorp, M., Zimmer, R., et al. (2005). *The role of districts in fostering instructional improvement: Lessons from three urban districts partnered with the Institute for Learning*. Santa Monica, CA: RAND Corporation.
- McLaughlin, M. W. (2006). Implementation research in education: Lessons learned, lingering questions, and new opportunities. In M. I. Honig (Ed.), *New directions in education policy implementation: Confronting complexity* (pp. 209–228). New York: State University of New York Press.
- McLaughlin, M. W., & Talbert, J. W. (2001). *Professional communities and the work of high school teaching*. Chicago: University of Chicago Press.
- Showers, B., & Joyce, B. (1996). The evolution of peer coaching. *Educational Leadership*, 53(6), 12–16.
- Slavin, R. E., & Madden, N. A. (1998). *Disseminating Success for All: Lessons for policy and practice* (Rev. Tech Rep.). Baltimore: Center for Policy Research on the Education of Students Placed at Risk, Johns Hopkins University.
- Slavin, R. E., Madden, N. A., & Datnow, A. (2005). *Research in, research out: The role of research in the development and scale-up of Success for All*. Baltimore: Johns Hopkins University.
- Success for All Foundation. (2005). *About SFAF—Mission, vision, and values*. Retrieved September 8, 2005, from <http://www.successforall.net/about/index.htm>
- Wilson, S., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. *Review of Research in Education*, 24, 173–209.
- Yin, R. K. (2003). *Case study research, third edition*. Thousand Oaks, CA: Sage.



# Assistance and Accountability in Externally Managed Schools: The Case of Edison Schools, Inc.

Julie Marsh and Laura Hamilton  
*RAND Corporation*

Brian Gill  
*Mathematica Policy Research*

Edison Schools, Inc., is the largest and most visible among a growing number of Education Management Organizations that have entered into contracts to manage public schools, including both conventional and charter schools. Edison's approach to managing schools is comprehensive, and it distinguishes itself from most other school improvement strategies by simultaneously addressing both the *resources and assistance* provided to schools and the *accountability systems* under which school staff operate. In this article we explore the ways in which the assistance and resources provided by Edison (including diverse professional development opportunities, materials, technology, and other tools), as well as accountability mechanisms (such as monitoring and rewards), have translated into principal and teacher actions, and the factors that facilitated or constrained educators' efforts to implement the Edison design and improve teaching and learning. Drawing on data gathered from extensive interviews, observations, and document reviews collected during a four-year comprehensive study of Edison schools, we demonstrate how Edison intends to promote not only educators' capacity but also their motivation and opportunity to deliver high-quality instruction. We examine variation that occurs across schools as teachers and principals respond to these system-level efforts. In addition, we identify several important predictors of variation in implementation, including the strength of instructional leadership provided by the principal and the presence or absence of district-imposed constraints such as union contract rules.

---

Portions of this article were drawn directly from a RAND report (see Gill et al., 2005).

Correspondence should be sent to Julie Marsh, RAND, 1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407. E-mail: [jmarsh@rand.org](mailto:jmarsh@rand.org)

New forms of governing and managing public schools have proliferated in recent years, and have led to rapid growth among companies that operate public schools under contract (Hentschke, Oschman, & Snell, 2003; Miron & Applegate, 2000). Many factors have supported this growth, most notably the proliferation of charter school legislation as well as accountability policies giving school districts the option and, in some cases, the mandate to contract out services for low-performing schools. Indeed, private management of public schools may continue to grow in the future, because the federal No Child Left Behind (NCLB) Act includes private management as one of the strategies that school districts may use to improve chronically low-performing schools. Both for-profit and nonprofit organizations have entered into management contracts with public schools, but much of controversy surrounding these Educational Management Organizations (EMOs) has focused on the for-profit providers. In the 2005–06 school year, for-profit EMOs were managing 521 public schools serving nearly 240,000 students across the United States (Molnar, Garcia, Bartlett, & O'Neill, 2006).

Among these EMOs, the largest and most visible is Edison Schools, Inc. In 2004–05 Edison served approximately 65,000 students in the schools it managed, and tens of thousands of additional students through other initiatives. Most of Edison's schools are operated under contract with local districts that have sought new management of existing schools, often because the schools have a long history of academic failure. Other Edison schools are brand-new start-ups, typically charter schools that Edison operates under contract with a local organization holding the charter, and Edison manages a few schools under contracts with states that have instituted takeovers as a result of chronic failure. Edison's approach to managing schools is comprehensive, and it distinguishes itself from most other whole-school reform strategies by simultaneously addressing both the *resources and assistance* provided to schools—such as professional development, materials, technology—and the *accountability systems* under which school staff operate, which include monitoring and rewards.

Because of Edison's prominence, it has been the focus of much of the debate surrounding for-profit EMOs. There has been limited empirical evidence to inform this debate. From 2000 to 2005 Edison contracted with RAND to conduct a comprehensive evaluation of achievement in Edison schools, and to examine Edison's design and how it is implemented in schools.<sup>1</sup> This article draws from the RAND study to describe Edison's approach to supporting school improvement, the ways in which the support strategies and accountability mechanisms translate into principal and teacher actions, and the factors influencing these efforts. We also present suggestive evidence of conditions that may influence achievement trends in Edison Schools. Specifically, we address three broad research questions:

---

<sup>1</sup>For further details, see Gill et al. (2005).

- What is Edison Schools' approach to and key strategies for supporting improvements in teaching and learning? What makes it different from other external assistance providers?
- How do these strategies play out at the school and classroom levels? What factors influence teachers' and principals' efforts to fully realize Edison's vision for improvement?
- What conditions and factors are associated with student achievement among Edison Schools? What school-level factors appear to matter most?

The answers to these questions illustrate the unique ways in which Edison has gone about supporting and scaling up teaching and learning improvements, and the factors influencing its efforts to translate its vision for a "world-class" education into a reality at the school and classroom levels. Of course, "scale-up" in the context of an organization privately contracted to run public schools means something different than it does with regard to other partnerships examined in this issue of the *Peabody Journal of Education*. Unlike support organizations that often seek to assist districts with improving teaching and learning in all of their schools, Edison's clients rarely want it to implement its model across an entire system. Instead, Edison may be one facet of a larger strategy to increase capacity to bring high-quality teaching to scale—for example, in Philadelphia, where Edison represented one partner in a larger "diverse provider" model in which many organizations received contracts to run various schools throughout the district. From the internal perspective of Edison Schools, scale-up also translates into efforts to enact its school design across a large number of schools throughout the country—a significant challenge of which is ensuring high-quality teaching and learning in a wide range of contexts and with support staff that are often not located in the same geographic area as the schools (i.e., a "virtual district").

In the following article, we first provide background on Edison Schools, including its history and past research on implementation and achievement, and describe our data sources and methods of analyses. We then describe Edison's overall approach to supporting improvement, followed by an analysis of how principals and teachers responded to these strategies in case study schools. Next we present a brief exploratory analysis of the relationship between implementation and achievement in our case study schools and conclude with implications for policy and practice.

## BACKGROUND ON EDISON

As one of the oldest EMOs in the country, Edison has spent more than a decade building its organization and system of schools. In 1991, Christopher Whittle, previous founder of Whittle Communications and Channel One News service for

schools, launched the Edison Project (renamed Edison Schools, Inc., in 1999). The Edison team spent three years developing a comprehensive school design that it regarded as exemplifying the best ideas from both education and business about curriculum, teaching methods, assessment, educational technology, staff development, and management. Edison sought to contract with school districts, charter-authorizing agencies, and charter holders to manage new and existing schools with this new design. Under these contracts, Edison would operate all aspects of the school, including curriculum, instruction, budgeting, hiring and firing, and staff development. The company would receive the same total average per-pupil funds available to local districts and “invest its capital up front on all new instructional materials, technology, and training to give the school a fresh start” (Chubb, 2004, p. 488).

In 1995, Edison opened its first four schools. For the next six years, the company experienced rapid growth, operating 133 schools by 2002. During this period of rapid growth, Edison leaders discovered that they needed systems to better support school design and implementation across a large number of schools (personal communication, 2002). Having spent much of the early history developing and refining the school design, Edison leaders built up new systems to better support and monitor operations and achievement. These systems, which have been refined over the years, are a major focus of this article.

After 2002, Edison’s expansion slowed amidst financial and political challenges, even as the company signed its largest single contract ever, to manage 20 schools in Philadelphia.<sup>2</sup> As of 2004–05, after several contracts were terminated for financial, academic, or political reasons, at the initiative of Edison or its clients, 103 schools were operating under Edison management. Edison continued to refine its system-level support for the schools it manages and began to diversify the portfolio of services it offers. In addition to its whole-school management partnerships with districts and charter authorizing agencies, the company offers other services such as its interim benchmark assessment system; technology and technical assistance with data; summer and after-school programs; supplemental educational services; and management consulting under the “Edison Alliance” flag, through which it offers access to many elements of its comprehensive reform model without taking on operational authority over a school. Although worthy of examination as other examples of external assistance to districts, this article does not focus on these other services but instead Edison’s whole-school management efforts.

In sum, in more than a decade of operating schools, Edison has gone from spectacular growth to retrenchment, a lower public profile, and diversification of its services. During this time, it also experienced an important shift in attention from

---

<sup>2</sup>For further details of Edison’s history in Philadelphia and the financial ups and downs, see Gill et al. (2005).



crafting its ideal school design to recognizing the need for developing system-level infrastructure and supports to ensure high-quality implementation of its design. These systems and supports are particularly important, given that implementation at scale has been a difficult problem for many school reform models (see, e.g., Berends, Kirby, Naftel, & McKelvey, 2001; Bodilly, 1998; Kirby, Berends, & Naftel, 2001).

## PRIOR RESEARCH ON EDISON SCHOOLS

Like comprehensive school reform models, Edison incorporates a broad set of services that are intended to be implemented at all of its schools—including a comprehensive curriculum package, enrichment programs such as foreign languages and art, instructional techniques, frequent assessments, professional development, extended school day and year, career ladders for teachers, and technology (discussed in more detail in subsequent sections). Unlike many comprehensive school reform models, however, relatively little research has examined achievement in Edison schools, and even fewer studies have investigated implementation of the Edison design.

The most comprehensive study of student achievement in Edison schools was completed by RAND in 2005 (Gill et al., 2005), which included both current and previously Edison-managed schools. RAND found that average gains in Edison schools during the first three years of Edison operation did not exceed the gains of matched comparison schools, but Edison results relative to comparison schools improved in years four and five. At that point, most Edison schools were matching or exceeding the gains of comparison schools, depending on the specific analysis conducted. One of the most important findings from that analysis is that performance may be a function of time: Edison schools' average performance improves as schools gain experience implementing the design.

As for implementation studies, the few that have been conducted suggest that although schools are able to enact many features of Edison's design, they vary in their ability to fully realize the ideal of the model (Government Accountability Office, 2002; Gomez & Shay, 2000; Rhim, 2002). For example, one evaluation of an Edison school found that it was able to implement several features with more fidelity—including the extended school day, extended school year, and daily professional development periods for teachers—but struggled with other features, such as achieving a "rich and challenging" curriculum, integrating technology, and implementing family partnerships (Rhim, 2002). Another single-school evaluation suggested that although the Edison design was well implemented, this implementation varied by year, growing stronger as the school remained under Edison's management over time (Gomez & Shay, 2000). In addition, this limited research identifies several factors affecting implementation. Some factors appeared

to constrain school efforts to implement the school design, including relationships with teachers' unions and teacher burnout and turnover because of rigorous demands required by the Edison design (Cookson, Embree, & Fahey, 2000; Rhim, 2002). Others facilitated fidelity to the Edison vision, including the investment in professional development opportunities for teachers, which enhanced teacher morale and enthusiasm (Cookson et al., 2000), and time (Gomez & Shay, 2000).

Building on this literature, our article seeks to understand Edison's overall model of improvement—with particular focus on the system-level resources, assistance, and accountability mechanisms—and the extent to which it translated into teacher and principal actions in a diverse sample of schools. Our findings add to the existing body of literature by examining how specific features played out in schools and classrooms and the challenges educators faced in enacting these features. It also adds exploratory evidence of school-level factors related to student achievement. The next section describes the data we examined and the methodology employed.

## DATA AND METHODS

This article draws on data collected from a variety of sources between 2000 and early 2005. The following section describes the sampling, data sources, and analyses we employed.

### School Sample

To examine school-level implementation of Edison's design, we visited 23<sup>3</sup> Edison schools that were selected to provide a range of school contexts and student populations. In particular, we selected schools to represent variation in local context (as represented by state and urban vs. suburban status), the year Edison began operating the schools (ranging from 1995, when Edison's first schools opened, to 2003), and the form of governance (i.e., charter schools and district contract schools).

---

<sup>3</sup>Not all 23 schools participated in the study from the start. We initially selected 15 schools in 2001. By 2003, three of the original 15 case study schools were no longer under Edison operation and a fourth elected to drop out of the study (and soon thereafter terminated its relationship with Edison). We replaced these four schools with four schools that were new to Edison, permitting us to maintain our sample size and a sample that better represented Edison's current portfolio of schools. In the fall of 2004, the RAND study team concluded that it would be useful to conduct a few additional school site visits as our study neared completion. Rather than return to schools we had previously visited, we elected to add four new elementary schools selected to add more balance in terms of governance and number of years under Edison's management.

TABLE 1  
Descriptive Characteristics of Case Study Schools and All Edison Schools

Schools	Case Study Sample (N = 23)	All Edison Schools Operating 1995–2005 (N = 144)
Charter school	43%	40%
Contract school	57%	60%
Start-up school	39%	31%
Conversion school	61%	69%
Opened 1995–97	26%	15%
Opened 1998–2000	48%	47%
Opened 2001–03	26%	35%
Located in Michigan	9%	14%
Located in Pennsylvania	17%	24%
Average total enrollment	581	662
Average % Asian	2	2
Average % Hispanic	16	21
Average % Black	60	62
Average % White	22	15
Average % FRL	70	74

*Note.* FRL = students eligible for free or reduced-price lunch.

Table 1 provides summary statistics on the sample of Edison schools, as compared with the full universe of Edison schools operating during the company's first decade. As the table indicates, the sample fairly represents the Edison universe on most key dimensions. The one respect in which the case study schools differ notably from the larger Edison universe is that, looking retrospectively at their full Edison histories to this point, their achievement results were somewhat better on average than those of other Edison schools, in both reading and math. Despite this average difference, however, the case study schools represent the full range of Edison's academic performance, with case study schools appearing in every quartile of the Edison-wide distribution of achievement trends.

## Data Sources

### *Case Study Data*

In 2001, 2003, and 2004 we observed classroom instruction and conducted extensive interviews with administrators, teachers, and staff in our sample of schools using semistructured protocols. In addition, we collected relevant documents (e.g., school improvement plans), listened to monthly "account review" calls in which Edison headquarters staff discussed our case study schools, and conducted telephone interviews of Edison regional staff responsible for overseeing our case study schools and relevant Edison clients (e.g., chartering authority officials, state and district officials).

### *Edison System-Level Data*

We conducted several rounds of interviews with Edison staff at all levels of the organization (corporate and regional offices) in 2000, 2002, 2004, and 2005, and we examined documents to understand Edison's strategies for school improvement and how these strategies translate into a concrete set of design components. We also observed several Edison conferences and professional development meetings in 2003 and 2004.

### *Student Achievement Data*

For the larger Edison study, we gathered student achievement data on state accountability tests for each Edison school and a set of comparison schools (see Gill et al., 2005, for details on the achievement data and selection of comparison schools).

## Analysis

Following each case study site visit, researchers analyzed all interview and observation notes and transcripts, as well as all documents collected on-site, and developed analytic memoranda summarizing overall findings about the school context and its implementation of the Edison school design. In addition, the RAND research team created a series of codes intended to measure the extent to which a wide range of design elements and contextual factors—from the implementation of each component of the curriculum, to the principal's skill as an instructional leader, to the existence of an extended school day and year—were present in each case study school. To ensure consistency, codes were assigned to each case study school during group meetings that included site visitors and other members of the research team. For most variables, codes were given in one of three categories: weak, moderate, or strong implementation. For some analytical purposes, we collapsed these into two categories: strong implementation versus anything less.

After coding all measures for each of the 23 case study schools, we combined several related measures into indices representing average results across several variables. Two of these indices figured prominently in our analyses. The first encompasses the implementation of curricula in subjects other than reading and math—that is, social studies, science, “specials” (art, music, physical education, and world languages), and core values (Edison's character education curriculum). These subjects constitute important elements of Edison's “world-class” educational model, but NCLB does not attach high stakes to test results in the subjects, so examining their implementation provides evidence of schools' attention to objectives other than those for which the state is holding them accountable. The



second index encompasses major features of what we characterize as the school's professional environment, including the use of houses, the availability of planning time, and the prevalence of site-based professional development. Both of these indices had high levels of internal consistency, as indicated by coefficient alpha estimates of .91 and .84, respectively.

We then divided the coded variables into two groups, largely corresponding to accountability systems and resources/assistance. We ran cross-tabulations and conducted exploratory statistical analyses that viewed the accountability measures as independent variables and the resource/assistance measures as dependent variables, permitting us to assess some of the underlying logic of Edison's strategies by examining relationships between accountability and resources. The aim was to examine in an exploratory way whether schools in which Edison's accountability systems are operating according to plan see better ground-level use of the resources/assistance in the Edison design. Where we found interesting and significant relationships, we report them throughout this article. In addition, by incorporating school-level achievement estimates for the case study schools, we were able to examine relationships between accountability systems and resources/assistance, on one hand, and student achievement outcomes, on the other.

## Limitations

Findings are based on examination of a relatively small number of Edison elementary schools that were not randomly selected. The aim of the case study examination was not to assess how Edison schools compare to conventional public schools (thus the absence of qualitative data collection from a non-Edison comparison group), but (a) to assess the extent to which Edison schools in practice match Edison's ideal in terms of design, and (b) to examine factors that might explain differences in school practices/teacher and principal actions and in student achievement among Edison schools. Given the limitations inherent in a small sample size, we sought a sample that would capture a wide range of Edison elementary schools to ensure sufficient variance in accountability systems, use of Edison resources and assistance, and achievement outcomes to permit us to understand how these various factors might be related. These analyses should be considered exploratory, and not necessarily generalizable to the full population of Edison schools.

The next section describes Edison's key strategies for supporting improvement, followed by a discussion of how these strategies played out at the school level.

## EDISON'S APPROACH TO SUPPORTING IMPROVEMENT

The stated educational aim of Edison's school-management business is the provision of "world-class education" to all of its students—defined as one that

“cultivates the mind to be ready for opportunities of every kind” in a rapidly changing world (Edison Schools, n.d.). In Edison’s view, this means that its students should have access to content in a wide range of subjects including arts and foreign languages. At the same time, Edison defines the critical primary measure of world-class education to be proficiency on annual state high-stakes assessments in reading and math (and additional subjects in some states). A focus on measurable progress on high-stakes tests has been a central characteristic of Edison since it opened its first schools more than a decade ago.

In the service of world-class education, Edison has devised a range of strategies to promote in its teachers and principals not only the capacity to deliver high-quality instruction but also the motivation and opportunity to do so. The attention to all three of these components—capacity, motivation, and opportunity—makes Edison’s strategies for student achievement unusually comprehensive. In Edison’s view

to change schools thoroughly, it is essential to change everything at once. Incremental reforms are too easily undone by those elements of the school that have not yet been changed. When everything changes at once, there are fewer old habits to break. (Edison Schools, 2004, p. 11)

Edison’s strategies for school improvement can be broadly classified into two categories: (a) Providing resources and assistance in support of a coherent and comprehensive school design, and (b) implementing accountability systems that aim to ensure that the resources and assistance for the design are in place and used as intended.

Figure 1 characterizes these strategies graphically. The resources and assistance Edison seeks to provide include technical capital (including curricula, assessment systems, and technology), human capital, social capital, and time, and they are directed at teachers, principals, and students and their families. Edison’s accountability model includes direct line and staffing authority, monitoring and rewards, parental involvement and market accountability, and the reduction of political and bureaucratic accountability that is prominent in conventional public schools. As we show in the next sections, Edison’s model is ambitious in its use of resources and assistance, but it is most clearly distinguished from conventional public schools (and from other providers of comprehensive school designs) in its accountability systems.

## Resources and Assistance

The resources and assistance that Edison seeks to put in place to build the capacity of teachers and principals are wide ranging, encompassing technical capital, human

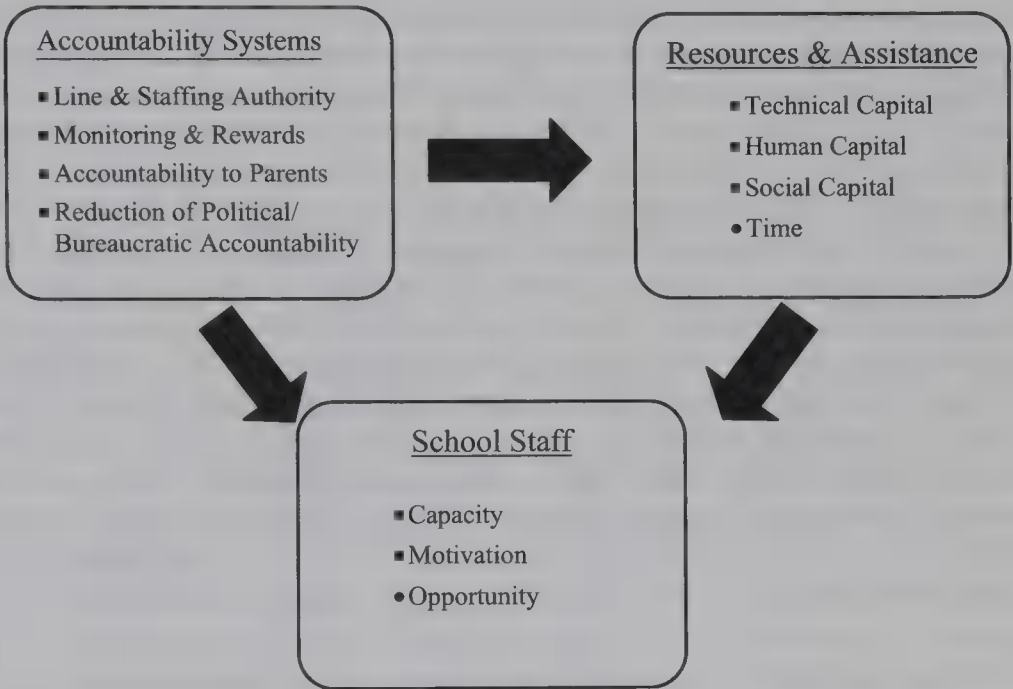


FIGURE 1 Edison's strategies for promoting school performance.

capital, social capital, and time. We briefly discuss Edison's vision for each of these components next.

**Technical Capital.** The key elements of the technical capital that Edison provides to schools include curriculum, assessments, and a variety of technology resources.

**Curriculum.** Edison's design teams selected programs they viewed as best supported by rigorous research (e.g., Everyday Mathematics in elementary grades) with some supplementary Edison-designed programs. Edison's curriculum goes beyond basic skills in reading and math to include explicit components in writing, social studies, science, art, music, world language, and fitness/health (Edison Schools, n.d.). Edison has sought to balance the need for standardization (considered essential for scaling up the model nationwide) and the need for flexibility (considered essential for promoting buy-in and adaptation to local norms as well as state-level policies; Chubb, 2004).

**Diagnostic assessments and analysis tools.** One of the key supports for the alignment of Edison's instructional programs with local standards and assessments is the Edison benchmark system, an online system of monthly

assessments developed by Edison. The benchmarks are monthly assessments in mathematics and reading, delivered online to Edison students in Grades two through eight, and are intended to provide rapid results and information to help teachers identify student needs and adjust instruction to meet those needs. The tests are customized to each state's standards and accountability tests, with items that resemble the content and format of the state tests. The benchmark results, therefore, can be used not only to diagnose the academic strengths and weaknesses of a student or a class but also to predict the likelihood that a student will achieve the state's proficiency standards. The system allows school staff to generate a series of reports that are designed to present information formatted in a user-friendly way. In recent years Edison has encouraged schools to use additional assessment data to guide instruction and instructional decisions, including the Dynamic Indicators of Basic Early Literacy Skills to gauge early elementary school student reading skills, and the Scholastic Reading Inventory to determine student reading levels. Edison regularly provides school staff hands-on training and tools to help them interpret and use achievement data.

*Other technology.* Telecommunications technology has represented a well-publicized part of Edison's academic model since the launch of its first schools, and although Edison has changed some aspects of its technology strategies, it continues to make substantial investments in technology in its schools. Teachers and administrators are given laptop computers, each classroom typically has a few desktop computers, and each school has a dedicated computer lab for communal use (in which benchmark assessments are administered, as well as instruction in computer skills). Edison has also created an intranet called The Common, a Web-based, "message, conferencing, and information system" that provides links to current research, curriculum materials, lesson plans, and discussion groups. Edison's technology investments also include telephones in every classroom and voice mail for every teacher, regarded as essential to promoting better, more-frequent, and more-efficient communication between teachers and parents.

### *Human and Social Capital*

Edison's support strategies aim to promote both human capital and social capital in its schools, addressing not only the capacity of its teachers and principals but also their motivation. These strategies include a variety of centrally provided professional development programs as well as school-site-based resources that are designed to develop teachers' knowledge and skills and to promote elements of social capital such as morale, trust, and school spirit.



*Centrally provided professional development.* Edison offers a wide range of professional development opportunities to its teachers and principals, beginning in the summer prior to the initial hiring of new staff (or the launch of the Edison school). These include the following:

- **Teaching Academies.** All teachers new to Edison are expected to attend intensive, week-long “Teaching Academies” delivered in large part by experienced Edison teachers who have been certified by Edison as trainers in the summer prior to their 1st year in an Edison school. The summer academies emphasize curriculum implementation, pedagogy, analysis of student achievement data, and classroom management. The summer academies also are used to begin building social capital, providing opportunities to establish relationships with teachers from other Edison schools, and using motivational programs that introduce teachers to Edison’s “culture of achievement.”<sup>4</sup>
- **Leadership training.** Edison principals and other school leaders receive approximately 2 weeks of leadership training each summer. The training focuses on analysis of data, specifically the use of the Edison benchmark system, as well as building management, improving curriculum and instruction, promoting staff capacity, supervision and evaluation, and the creation of a strong school culture. As with the teacher academies, this training is intended to promote social capital as well as human capital.
- **Achievement Academies.** Edison conducts regional “Achievement Academies” during the fall, aimed primarily at principals and school-level curriculum coordinators (teachers responsible for coordinating site-level implementation and professional development (PD) for a particular subject). These academies provide strategies that will enable schools to increase achievement for all students, as well as time that is reserved for work sessions, in which school teams utilize the strategies to update and revise their own individual School Achievement Plan.
- **Principals’ Leadership Conference.** Each fall Edison gathers its principals in a Principals’ Leadership Conference (PLC), at which it provides additional leadership training and offers recognition to the principals of high-performing schools.
- **Edison Evenings.** In recent years, Edison has begun offering an ongoing series of small-dose professional development opportunities in the form of “Edison Evenings,” voluntary training sessions on particular topics, delivered

---

<sup>4</sup>This includes methods to increase student motivation to achieve (e.g., displaying exemplary work), to involve parents in supporting the school (e.g., advisory councils), and to recognize and reward staff for performance. See Gill et al. (2005) for further discussion.

via conference call and computer linkup to interested Edison teachers across the country.

*Ongoing support from Edison staff.* In each subject area, Edison maintains a small staff of full-time curriculum experts (often individuals who previously taught in Edison schools), known as National Curriculum Coordinators, who aim to provide systemwide professional development and support individual schools as needed, via e-mail, telephone, and occasional site visits. The schools' primary contact for instructional support purposes is an Edison regional Achievement Vice President—an individual assigned to about seven schools who provides support and assistance to principals and school staff on all matters related to instruction and student achievement, as well as design implementation and student discipline. The Achievement VPs—who, like the National Curriculum Coordinators, are usually former school-level staff, either principals or teachers—assist schools in making plans for student achievement, in analyzing test results, in complying with the demands of NCLB and state accountability systems, and in executing basic program components (e.g., ensuring that curriculum coordinators develop observation schedules or that newly hired teachers attend training).

*Supports to develop site capacity and school-based professional development.* Edison's approach to professional development includes a variety of day-to-day activities that are expected to occur at the school site and are usually led by school staff. In addition to the training previously described, Edison seeks to develop site capacity in several ways:

- **Standards and rubrics for instructional leadership.** Edison leaders believe that principals should be instructional leaders, as well as good managers of building and budget, and facilitators of a strong school culture to ensure results in five key areas—student performance, school design, customer satisfaction, financial management, and operational excellence. In support of this view, Edison has developed detailed standards and rubrics specifying principal expectations, which are used in an annual appraisal process.
- **Distributed leadership model and roles.** The Edison school design tries to distribute instructional leadership responsibilities and capacity among teacher leaders in the school. Each school is supposed to have a leadership team that is responsible for helping the principal develop, adjust, and monitor school policies, procedures, and programs. The leadership team includes not only the principal and the academy directors but also the lead teachers for each of several small “houses” into which the school is organized. Each house consists of about six teachers, usually representing two or three grade levels (within an “academy”). Students are expected to stay in the same

houses as they progress through several grades, so that the team of teachers in the house can be responsible for instructing and managing a common group of students over time. The lead teacher for each house is expected to serve as a mentor for the other teachers with respect to both pedagogy and classroom management, and (where permitted by contract) to take some responsibility for the evaluation of junior colleagues within the house. Each house team is expected to meet daily, and each daily meeting is intended to be an opportunity for teachers to work together and develop their skills (Chubb, 2004).

Each subject area has a curriculum coordinator in the school who is generally a teacher given additional responsibilities including managing curriculum materials, providing ongoing professional development in the curricular area, conducting classroom observations and modeling instruction. Site capacity under the Edison design also includes full-time staff who are responsible for the school's special education program and for student and family support related to behavioral challenges and special needs.

### *Time*

Edison describes "A Better Use Of Time" as one of its key strategies to promote student learning (Edison Schools, n.d.). This involves, first of all, a substantial increase in total instructional time for students. Ideally, the standard Edison school year is expected to be 198 days, about 10% longer than the 180 days required in most states. The standard Edison school day is expected to be longer as well, an hour or more beyond the time expected of most public-school students. The additional time is intended to help fit in all components of the curriculum and provide teachers with two periods a day for planning and professional development. The "better use of time" also involves the creation of a "safe and orderly learning environment" (Edison Schools, 2002a, p. 5) that is intended to allow teachers to focus on teaching—as opposed to discipline problems and other related issues that detract teachers' time and attention away from instruction. To support such an environment, Edison-developed "character and ethics" curriculum promotes the teaching and modeling of core values—wisdom, justice, courage, compassion, hope, respect, responsibility, and integrity—throughout the school.

### *Accountability Systems*

What makes Edison, like other EMOs, novel on the American K-12 education scene are the accountability systems it intends to establish both within its schools and across its system. Unlike other providers of educational services and comprehensive school reform models examined in this issue, EMOs have operational

authority over the schools in which they work. Edison seeks to use this operational authority to impose accountability systems that supplement or replace many of the conventional accountability systems of American public schools. Edison's accountability model begins with straightforward line and staffing authority, adds a system of monitoring and rewards, and includes the reduction of conventional political and bureaucratic authority.

### *Line Authority*

Edison regards operational authority over its schools as crucial. Principals in Edison schools report to regional "general managers," who in turn report to Edison executives in the New York headquarters. Edison's chief education officer serves a role analogous to that of a school district's chief academic officer and Edison's CEO is much like a district superintendent.

### *Staffing Authority*

In Edison's view, one of the key aspects of operational authority over schools is the ability to hire and fire staff. Staffing authority, according to the Edison model, is important not only for ensuring the effective operation of line authority but also for promoting the buy-in of staff. Because Edison's school design is demanding and highly specified, it is especially important that its principals and teachers are supportive; voluntary transfer in and out makes that support more likely. Authority over staffing involves more than just hiring and dismissal. Edison has developed a career ladder internally that aims to give teachers opportunities to advance to greater responsibility and salary, in positions such as school-level curriculum coordinators and house lead teachers, without leaving the classroom for administration and on the basis of competence not seniority.

### *Monitoring and Rewards*

Edison attends to the motivation of its staff not only with opportunities for advancement but also with systems to monitor and reward performance.

*Information collection systems.* Edison utilizes multiple means to gather information on design implementation, instructional performance, and student achievement in its schools. These include in-person visits to schools, monthly calls in which Achievement VPs and other corporate staff discuss each school's progress, and reviews of schools' Benchmark Assessment data.



*Star rating system.* Edison's "star rating" system is its key instrument for determining a school's eligibility for performance-based rewards. The system is designed to be "an objective measure from which we can celebrate success or set targets for improvement" (Edison Schools, 2002b, p. 3). Each year, Edison rates each of its schools in terms of five characteristics or "Points of Accountability," which it defined for its principals at their 2004 leadership conference as follows:

- **Operational Excellence** measures "the factors that we know are keys to healthy and successful schools," including student attendance, staff attendance, student mobility, teacher turnover, and graduation rate.
- **Customer Satisfaction** measures "a school's ability to please its students, parents, and staff" and averages student, parent, and staff ratings from surveys to determine overall customer satisfaction.
- **School Design** measures implementation of the "Edison Ten Fundamentals," including school organization, use of time, curricular program, instruction and pedagogy, assessment and accountability, professional development, technology, partnership with families, communications and community outreach, and system growth.
- **Financial Management** measures the fiscal health of the school and is determined in multiple ways, depending on the nature of Edison's contract with its client. Usually, successful financial management in an Edison school requires the school to meet an enrollment target.
- **Student Achievement** measures student learning and is determined by a complex formula that emphasizes relative growth in schoolwide proficiency rates as measured by state-mandated tests—and, more recently, by the ability to meet Adequate Yearly Progress.

Edison staff have developed detailed criteria and rubrics for awarding each school one to four "star" ratings in each of the five areas. Edison uses the star rating system to recognize and reward school and individual performance. Where allowed by contract, principals and teachers are also eligible for monetary bonuses based on weighted star ratings, which primarily emphasize student achievement and factors tied to academic success.

### *Other Accountability Mechanisms*

In addition to these formal school-based accountability elements, Edison schools differ from most other public schools in their accountability to parents, which is achieved through choice-based assignment, parent advisory councils at each school, parent satisfaction surveys, and requirements for parents to attend quarterly conferences with their children's teachers. Edison tries to reduce bureaucratic accountability by giving principals more authority over budgeting than they

would have in conventional public schools—and, as a corollary, more freedom from the bureaucratic constraints that are typically imposed by districts. Edison also aims to insulate its schools from local politics, in the hope that this will maximize opportunities to focus on instruction. This aspect of Edison's accountability strategy is derived directly from the insights expressed in *Politics, Markets, and America's Schools*, in which Chubb and Moe (1990) argued that the direct operation of public schools by elected officials frequently prevents them from focusing intensely on their academic missions (see also F. M. Hess, 1999; Hill, Pierce, & Guthrie, 1997).

### Edison Strategy Summary

In sum, the assistance and accountability systems that constitute Edison's strategies for promoting student achievement are intended to address all elements relevant to high-quality delivery of instruction, including capacities, motivation, and opportunities for school staff. In the next section we explore the extent to which Edison's strategies are realized in practice in a sample of its schools.

## HOW EDISON'S IMPROVEMENT STRATEGIES ARE REALIZED IN PRACTICE

This section examines the ways in which teachers and principals responded to Edison assistance and accountability mechanisms. As we describe, nearly all of the Edison schools we visited across the country showed enough consistency of implementation to be clearly recognized as Edison schools, but we observed considerable variation in the extent to which they fully realized the Edison ideal. We start by examining how educators responded to Edison's accountability systems, followed by an analysis of their responses to the key assistance mechanisms.

### Accountability Systems

As Edison leaders have acknowledged, they do not always have the opportunity to fully implement all of the accountability systems that their design involves. Each of Edison's contracts to operate schools is unique, and clients sometimes impose constraints that require compromises to Edison's ideal model.

#### *Line Authority*

As intended, Edison had operational authority over all of the case study schools we visited, with principals reporting to Edison's regional general managers. But Edison's authority over school operations was not always complete, and principals in some district partnership schools complained of the challenges associated with

reporting to “two masters”: Edison and the district. Edison’s charter schools usually had fewer problems with competing authority, but local charter boards sometimes sought to assert their influence, occasionally creating challenges similar to those experienced in many district schools. The ability to navigate the political and contractual waters associated with having two masters was a critical skill both for Edison principals and for general managers responsible for maintaining good client relations. In extreme cases, district clients viewed Edison as a mere vendor—providing curriculum, professional development, and assessment tools—rather than a manager with both the responsibility and the authority to run the school. Our case studies included a small number of schools where district clients had this attitude, and such schools typically only weakly represented the Edison culture.

### *Staffing Authority*

Along with operational authority, Edison had authority to hire and fire the principal in nearly all of the schools we visited. Edison sets high expectations for principals, and it had dismissed more than a few who had fallen short. On at least one occasion it set a target of improving or dismissing the bottom quartile of principals, and followed through on the plan, firing 80% of the bottom-quartile group. In 2004–05, Edison made a point of evaluating principals early in the year, and dismissed at least two in midyear.

By contrast, we observed a few schools in which Edison’s nominal authority over the staffing of the principal position was undermined in practice by the principals’ personal relationships with the clients (district or charter authorizer staff). In short, Edison’s *de facto* authority to dismiss a principal is sometimes less than the letter of the contract might imply.

The authority to dismiss an ineffective principal appears to matter. Edison case study schools in which RAND researchers gave principals strong ratings for instructional leadership (i.e., principals who appeared to spend a substantial amount of time visiting classrooms, who analyzed achievement data, and who took an active role in site-based professional development for teachers) also showed stronger implementation of both tested (reading and math) and nontested (science, social studies, specials, and core values) aspects of the Edison curriculum.<sup>5</sup> Moreover, schools with strong instructional principals had better achievement results (as we discuss further at the end of this article).

---

<sup>5</sup>On a reading/math implementation scale ranging from one to two, Edison schools with strong instructional leaders had a mean score of 1.89, whereas schools without strong instructional leaders had a mean score of 1.61 ( $N = 18$ ). On a nontested subjects implementation scale ranging from one to two, Edison schools with strong instructional leaders had a mean score of 1.68, while schools without strong instructional leaders had a mean score of 1.29 ( $N = 18$ ). In both cases, differences were statistically significant at  $p < .05$ .

Edison's authority over *teacher* staffing was more often compromised than its authority over principals, largely because it usually was required to honor existing teacher contracts in its district partnership schools. In most of Edison's charter schools, teachers were employed under one-year contracts that were renewed at the discretion of the principal. But in district schools, Edison teachers usually had the same contractual and tenure protections as teachers in other public schools in the local district. (And we often heard principals in district contract schools long for the staffing authority available to charter school principals.) Edison teachers in district schools often received their paychecks from the district rather than from Edison.

In general, Edison was willing to accept compromises to its ability to dismiss teachers as long as the district made it relatively easy for teachers to voluntarily transfer out of the Edison school (see Chubb, 2004). Edison leaders believed that, in most instances, the voluntary transfer mechanism would ensure that the teachers who do not "buy in" to the Edison model would not stay. Consistent with this view, we saw only one Edison school that had substantial numbers of teachers who were actively opposed to Edison.

Within each Edison school we visited, the assignment of teachers to leadership positions—that is, the use of Edison's teacher career ladder—at least nominally followed the Edison design. Principals had the authority to appoint lead teachers and subject-matter curriculum coordinators in the case study schools, and they were not required to abide by seniority rules in making such appointments. In many district partnership schools, however, existing teacher labor contracts constrained Edison's ability to set salaries commensurate with the teacher ladder (rather than with seniority). Many of the young teachers we spoke with (and Edison's teachers are often young) looked favorably on these leadership opportunities, even if those opportunities did not include substantial pay benefits. They appreciated the chance to assume positions of instructional leadership in the school, earlier than would be possible under a seniority system. For instance, a lead teacher told us the career ladder provided teachers with the "incentive to strive, to be there."

### *Monitoring and Rewards*

We observed a wide range of responses to Edison's systems for collecting information and rewarding schools and staff. First, we found a high level of detail in the conversations among Edison's central and regional staff occurring during Edison's monthly account review calls, which suggested an understanding by Edison staff of principals' instructional leadership capabilities, of the general quality of instruction in the school (particularly as related to subjects that are included in state assessments), and of the strengths and weaknesses of teachers. The integration of monthly test results and qualitative assessments by direct observers added



to the quality of these conversations. All of this information permitted Edison staff on the calls to develop targeted strategies to address problems that came up.

Nevertheless, Edison's systems for monitoring achievement and instruction had some weaknesses, driven in some cases by economics and geography. Even though each Achievement VP was typically responsible for only seven schools, those schools were in some instances widely dispersed geographically, making it difficult for the Achievement VPs to visit regularly. Moreover, Edison's information about staffing in schools was often unreliable, because data systems for staffing very often ran through the local school district rather than through Edison.

Second, in the schools we visited, Edison's star rating system had substantial success in getting the attention of principals and mixed success in getting the attention of teachers. This difference is related to the fact that substantial bonuses tied to star ratings were available to most (but not quite all) Edison principals, whereas contracts often precluded bonuses being given to teachers. Moreover, even where teacher bonuses were available, the bonus pool depended on the performance of the entire school rather than individual teachers. Within a school, the distribution of bonuses among teachers was typically at the discretion of the principal.

A minority of principals we interviewed expressed frustration at the complexity of the star rating system, perceiving it as mysterious, arbitrary, and at least partly beyond their control. More often, principals reported that their own motivations were primarily intrinsic, but that the availability of bonuses was a nice benefit. As one principal told us,

I don't really think that, if a principal gets up everyday, a bonus is what they're truly after. It's a nice ending to a year of hard work, but I don't think that's what really pushes them to reach that. I think it's the children.

To the extent that the star rating system motivated behavior in the schools, it was reinforcing the same signals that are created by NCLB and attendant state high-stakes testing systems. We observed an intense focus on achievement on state accountability tests in many of the Edison case study schools—leading to practices both consistent and inconsistent with the Edison ideal of a “world-class” education that is both broad and deep. As we discuss later in this section, in some instances, a focus on test scores undermined the commitment to nontested subjects. Another consequence of NCLB that we increasingly observed in Edison central office discussions and in schools during the course of our study was a focus on “bubble kids”—that is, students whose current achievement levels place them near the state's cutoff for determining proficiency in reading and math. In response to NCLB, which requires all states to establish school accountability systems based on the proportion of students achieving proficiency and which sanctions schools and districts based on these proficiency targets, many public schools around the

country have sought to identify and direct interventions toward those students who are closest to the cut-point for proficiency (Booher-Jennings, 2005; Hamilton et al., 2007; Pedulla et al., 2003). Edison's monthly benchmark assessments gave its schools unusually good information for identifying bubble kids, and Edison actively encouraged schools to identify such students and develop interventions to prepare them for state exams.

In the Edison schools we visited, there was some variation in attention to bubble kids. Some Edison principals and teachers embraced the concept as a logical and appropriate way to have data drive instructional decision making. Others, however, were disturbed by the possible implication that students on both ends of the achievement spectrum—high achievers and low achievers—might be neglected in favor of those in the middle. These educators tried to maintain an instructional focus on improving the achievement all of the children in their schools, regardless of their current proficiency levels.

### *Other Accountability Mechanisms*

Although site visitors had little opportunity to observe the case study schools' interactions with parents, our conversations with Edison teachers provided one indication that the communication was occurring. In nearly every Edison school we visited, teachers reported high levels of parent participation (typically better than 90%) in quarterly report card meetings. Edison's requirement that its report cards be given to parents in person appeared to be effective in bringing them to the school several times a year to meet with teachers.

As for being "schools of choice" as intended, the extent to which parents and students actively chose Edison schools varied considerably across our sample. Enrollment in an Edison charter school usually required an active choice by the family. Although also true in some Edison's district contract schools, others retained neighborhood assignment schemes in which parents had to actively opt out if they wanted their children to go to school elsewhere. Interestingly, we did not observe substantial differences between charter schools and district contract schools in terms of the implementation of Edison's curricula or of elements of the school's professional environment (i.e., houses, planning time, and site-based PD).

Edison's effort to clear away some of the bureaucratic constraints on its principals had only mixed success in the case study schools we visited. Many of the case study principals had greater authority over school budgets than they would in conventional public schools, but this authority varied widely, depending on the particular contract that Edison had with its client. Edison principals who were constrained by district requirements sometimes expressed frustration that they lacked the authority available to their colleagues, particularly in charter schools. Principals in district contract schools more often had to deal with external bureaucratic

challenges, related to issues such as building maintenance, budgets, paperwork, materials, or district-sponsored professional development.

The additional local bureaucratic and contextual obstacles that some Edison schools faced may have affected the implementation of the design. In our case study sample, schools where staff reported more local constraints had weaker results on the professional environment index, suggesting more difficulty in implementing the Edison house structure, the planning periods, and site-based professional development.<sup>6</sup> In some schools, for example, Edison was unable to implement its longer school day, which in turn prevented the implementation of its standard of two daily planning periods for teachers. Across our case study schools, however, we did not observe a relationship between local constraints and the implementation of the curriculum.

### Assistance and Resources

We now turn to the assistance provided to Edison schools and the responses we observed within case study schools.

#### *Technical Capital: Curriculum*

Nearly all of the Edison schools we visited, in all parts of the country, were immediately recognizable as Edison schools, by virtue of the curriculum materials and examples of student work covering nearly every wall, in classrooms and hallways alike. Only two of the case study schools demonstrably deviated from the standard Edison appearance, and in those two schools the absence of Edison wall displays was a clear sign of much deeper problems with the commitment of the staff to the Edison model.<sup>7</sup> The various materials associated with the Edison curricula (including textbooks and manipulatives) were consistently present in the schools we visited, although many schools reported delays in receiving the materials during their first year of operation.

Our teacher interviews and classroom observations provided only a limited view of the implementation of the curriculum in the classroom. Not surprisingly, there appeared to be more implementation challenges during the first year of Edison operation than during later years. Many of our study participants said that learning

---

<sup>6</sup>Schools without substantial local constraints had a mean score of 1.77 on the professional environment index (for which scores ranged from one to two), whereas schools with substantial local constraints had a mean score of 1.57. The difference is statistically significant at  $p < .05$ .

<sup>7</sup>One of these schools was a very troubled 1st-year Edison school in Philadelphia, whereas the other was a long-time Edison school, which not long after our visit ended its contract. Both schools had serious problems with leadership and morale.

how to teach using the new programs was difficult, particularly given that new curricula in every subject were being introduced simultaneously. In most of the schools we visited, implementation appeared to be strongest in reading (Success For All (SFA) or Open Court) and math (Everyday Mathematics)—consistent with the emphasis of Edison’s central office, and with the incentives created by most states’ test-based accountability systems. Nearly every school followed a schoolwide daily schedule involving 90 minutes of simultaneous, mandated reading instruction for all students, and at least 60 minutes of daily mathematics.

In some schools we were told of occasional displacements of curriculum altogether, but we saw no evidence that this occurred frequently. More often, we learned of schools supplementing the curriculum with additional materials designed to prepare students for state exams. Edison’s flexibility in allowing schools to supplement the curriculum to meet the needs of local standards and assessments and its efforts to provide teachers with tools to embed test skills within and alongside the existing curriculum, as well as the time available in the long school day, may have contributed to maintaining the fidelity of implementation of its core programs in reading and math.

By contrast, we heard more reports that “nontested” elements of the Edison curriculum were sometimes displaced by test preparation or other priorities. Implementation of Edison’s curricula in social studies, science, and “specials” (including art, music, and foreign language), was less consistent than the implementation of the reading and math curricula across the case study schools. A few teachers suggested that this displacement resulted in part from Edison’s own focus on reading and math. External pressure from states’ test-based accountability systems (which usually focus on reading and math) undoubtedly contributed as well, as it does in other public schools. Compromises in the implementation of nontested subjects were in some case study schools related to resource limitations. In Philadelphia, for example, Edison’s contract with the district did not provide sufficient resources to fully implement the model, forcing the abandonment of the longer day, the longer year, and some of the fine arts curricula. According to Edison central office interviews, the budget crises that hit states and local governments across the country in the early part of this decade led to similar compromises in many of its schools.

### *Technical Capital: Diagnostic Assessments*

Early in the development of the benchmark system, we observed a variety of implementation challenges in the schools. Benchmarks were originally issued on paper, which meant they required time to assess. The launch of the electronic benchmark system was plagued by a variety of technical problems, leading to frequent frustration in many schools when the system was overwhelmed. By the time of our second round of visits, however, these problems had been largely



ironed out, and the system appeared to be used faithfully and reliably at nearly all of the case study schools (with the exception of some new start-up schools).

More importantly, reports indicated that many teachers and principals found the benchmarks valuable, and were using the results effectively and as intended to diagnose instructional challenges and develop interventions. Reports of the misuse of benchmarks (for example, interpreting them as high-stakes assessments and providing preparation specifically for benchmark tests) were rare in the case study schools and were vigorously countered by clear messages from Edison's central office about appropriate use. Based on conversations with school administrators at the Principals' Leadership Conference, the provision of additional diagnostic instruments (such as Dynamic Indicators of Basic Early Literacy Skills) and analytic tools were also valued and much used.

### *Technical Capital: Technology*

Our case study schools provided a few examples in which Edison's investments in computers and audiovisual technology were being well used by students as well as teachers, for example, in conducting a daily student-run live video announcement delivered to all classrooms at the beginning of the day. By 2003, virtually all of the case study schools were participating in the monthly online benchmark assessments. But with the important exception of the benchmarks (discussed previously), we saw little evidence of a systematic, Edison-wide plan for the use of computer technology in the curriculum, despite some investments such as state-of-the-art experimental computer labs installed at a couple of schools. Moreover, school staff frequently complained to us about technical problems, especially in the first year of the school's operation, and insufficient technical support from Edison. The schools that were making extensive use of computers in instruction appeared to be doing so largely at local initiative. We have no reason to believe that Edison schools are trailing other public schools in the use of computers in instruction, but the reality in the Edison schools is well short of the high expectations that Edison created for its clients. As of this writing, however, Edison has launched a major research and development project that is, among other things, preparing to substantially increase the role of instructional technology in Edison Schools.

Unlike instruction, communication in Edison schools was clearly advanced by Edison's technology investments. Teachers generally appreciated Edison's provision of laptops and e-mail (except in the few cases where budget constraints precluded the provision of laptops), and many teachers used them to correspond both with colleagues in the school and with Edison's regional and national staff. Edison's intranet, known as The Common, was used less consistently but was regarded as an asset by the teachers and principals who took advantage of its resources. Many teachers also noted that these investments benefited

parent communication. The phones and voicemail made it easier for parents and teachers to communicate about homework assignments and behavioral challenges. Teachers also valued them as an indication of professional respect, so the investment actually aided teacher morale in some schools.

### *Time*

Most of the case study schools we visited used an extended school day (19 of 23 schools) and an extended school year (15 of 23 schools), as intended in the Edison design. Edison's Philadelphia schools did not operate with a longer day and year, as a result of contractual and resource limitations there. Outside of Philadelphia, some Edison schools had shortened their academic year, in part because of resource limitations and in part because of concerns about teacher burnout. Edison leaders believed they had been more successful with the longer school day than with the longer year, for a number of reasons. Attendance was usually lower during the additional weeks of school, because families may have had other children in schools using conventional calendars and therefore may have been unprepared to have their children in school, and state attendance requirements could create unintended problems for schools with a longer year, if attendance was measured during those weeks. Despite these challenges, many Edison schools not only maintained longer standard schedules but also operated after school and Saturday programs to provide additional skill training, especially for bubble kids and especially in the weeks prior to state exams. Finally, most of the case study schools were able to put in place the two periods of daily planning and professional development as intended, but some (such as those in Philadelphia) had difficulties related to local contractual issues.

With respect to the quality of classroom instructional time, site visitors observed in classrooms of case study schools across the country teachers using various classroom management techniques that Edison taught to all teachers in the Teaching Academy, and they appeared to be effective in keeping students focused and alert, and maintaining a "safe and orderly learning environment." In addition, teachers in many schools made effective use of the house support structure to handle behavior problems before they required the attention of the school administration. In a few schools with serious and chronic discipline problems, these appeared to be associated with weak building management on the part of principals.

### *Human and Social Capital*

We now turn to Edison's investments in the skills, morale, and trust of its school staff, addressing first the professional development resources provided by Edison's central office, and then the school-site mechanisms for professional development.

*Centrally provided professional development.* Edison's up-front investment in the skills of teachers and principals was generally well received. Edison teachers often described the summer Teaching Academy as overwhelming in its intensity and the breadth of its content, and they reported that the value of the seminars varied with the skills of the presenter. Nevertheless, they were typically pleased with the simple fact that Edison paid for their participation in a weeklong conference at an out-of-town hotel. The professional development conferences were also viewed by many Edison teachers as a sign that they were respected as professionals. This had benefits in terms of morale and trust even apart from the substantive benefits the training may have had for the skills of teachers.

Perhaps the primary concern about Edison's professional development conferences for teachers was that the investment was often lost as a result of attrition. Edison-wide, rates of teacher attrition were unclear,<sup>8</sup> but it was a serious challenge at many of the case study schools—as it is at high-poverty urban public schools generally, which, like Edison, rely extensively on early-career teachers who have the highest rates of departure from the profession. Edison recognized that its efforts to build site capacity in schools were often hampered by attrition, and it had incorporated retention of teachers into its star-rating formula in an effort to encourage principals to promote stability.

Like teachers, many principals valued the summer leadership academy and the fall PLC as indicators of professional respect. We spoke to a number of Edison principals who appreciated the responsibility and support that Edison provided them, particularly in the area of instructional leadership. Relatively new Edison principals were pleased not only to be attending the conferences as learners but also to be given the opportunity to present to their colleagues. Nevertheless, like new Edison teachers, many new Edison principals found the experience overwhelming, particularly if they did not have prior experience acting as instructional leaders or managing budgets. A number of new principals told us they would like more support from Edison in these areas. Given the high expectations that Edison had for principals, and the extensive demands it placed on them, the PLC and its annual awards ceremony were particularly important for promoting morale and a sense of Edison-wide community among principals (although we heard some disgruntlement from a small number of principals who felt that unfair financial targets made it impossible for them to win awards).

*Ongoing support from Edison staff.* Because of Edison's reorganization that reduced its reliance on in-person school visits from its national curriculum

---

<sup>8</sup>Edison schools typically reported some sort of teacher turnover rate, but reported rates were based on local definitions of turnover and were therefore not necessarily comparable across Edison schools. We were unable to calculate an Edison-wide teacher turnover rate with confidence.

staff in favor of greater reliance on the regional Achievement VPs, the number of support visits to Edison schools appeared to decline, as Edison tried to reduce travel costs incurred by its central staff. The curriculum staff tried to replace some of the reduced school visits with remote support, via e-mail and phone, and via regularly scheduled “Edison Evening” professional development programs conducted by conference call. School staff generally appreciated their e-mail and telephone access to Edison’s national curriculum staff, but many of them would have liked more in-person support. We heard more complaints about insufficient support in schools that were relatively isolated geographically (unlike those in Philadelphia, where staff felt well supported). For example, one staffer at a relatively new, and struggling, Edison school complained that “I feel like we’ve been left in the lurch.” School staff members who interacted with Edison’s national support staff were usually pleased with the quality of the support, although many of them would have liked it in greater quantity. We heard some complaints, however, in areas like science, where Edison invested fewer resources than in math or reading, and where many states did not yet have accountability tests.

*Supports to develop site capacity and school-based professional development.* Across schools, we saw wide variation in the extent to which the teaching staff viewed the principal as an effective instructional leader. Some Edison principals focused on the more traditional responsibilities associated with building management. Others, however, appeared to be highly successful at leading training sessions, modeling instruction, and motivating teachers. In our case study sample, there was some evidence that charter schools were more likely to have strong instructional principals than were district schools.<sup>9</sup> We can only speculate on the reason for this, but it may be related to the fact that charter schools were less likely to be bound by teacher contracts that narrowly define the scope of a principal’s instructional supervision responsibilities.

As previously noted, strong instructional leadership by principals in the case study schools was associated with stronger implementation of both tested curricula (reading and math) and nontested curricula (science, social studies, specials, and core values). Moreover, we also found that Edison schools with weaker instructional leaders were more likely to subsequently end their contractual relationship with Edison than were schools with strong instructional leaders.<sup>10</sup>

---

<sup>9</sup>Five of eight charter schools in the sample were coded as having strong instructional principals, whereas only two of ten district schools were rated with strong instructional principals. (In five schools, we lacked sufficient information to make a judgment about instructional leadership.)

<sup>10</sup>In the small sample of case study schools for which we were able to rate instructional leadership, only one of seven (14%) schools that later ended relationships with Edison had strong instructional leaders, whereas six of 11 (55%) schools that remained with Edison had strong instructional leaders.



The extent to which Edison's ideal of distributed leadership was implemented in the case study schools varied widely. In the best-functioning schools, the teaching staff viewed the opportunity to participate in schoolwide decisions via the school leadership team as one of the best features of the Edison design. In such schools, lead teachers appreciated the empowerment represented by participation in the leadership team. As one lead teacher noted, "To have ownership in something, you need to feel [you are a] part of it." The extent to which the leadership team involved genuine collaboration in decision making depended almost entirely on the personal style of the principal; some welcomed shared leadership, whereas others preferred a more autocratic model. We did not observe that this difference predicted a school's achievement results.

Similarly, the house structure was formally present in virtually every Edison school we visited, but its effectiveness and use depended on the skills and ambition of the lead teachers. We encountered examples of houses in which lead teachers provided active mentorship to their junior colleagues, assisted with behavior problems in the classrooms of other teachers in the house, and played an active role on the school leadership team. By contrast, some lead teachers lacked the capacity, motivation, or respect from their house members that would have been needed to take on the leadership responsibilities. In some schools, particularly in the start-up year, principals had difficulty finding experienced and motivated teachers to take on the role. Apart from the training that Edison provided them, few lead teachers had any prior training or experience in the evaluative role of the lead teacher—a role that represented a substantial cultural shift for nearly all teachers. In many district contract schools, however, this problem was rendered moot, because lead teachers were prohibited by the teacher contract from serving as evaluators.

Some of the challenges facing both lead teachers and school-level curriculum coordinators were inherent in Edison's model. In particular, although Edison sought to give substantial instructional leadership responsibilities to lead teachers and coordinators, the model did not give them additional time during the school day to pursue those responsibilities. In schools using SFA, the reading coordinator was freed of teaching responsibilities for half of the day; in other subjects, coordinators were full-time teachers who were expected to fulfill their responsibilities during their standard PD periods and when principals could find occasional substitutes enabling them to observe the instruction of their colleagues and provide coaching support. Curriculum coordinators in Edison schools across the country told us that they rarely had opportunities to get out of their own classrooms and act as coaches—a key task of a curriculum coordinator, according to the Edison design. As a result, many curriculum coordinators defined their jobs largely in terms of keeping track of the inventory of materials for their subject matter.

## Summary

The best-functioning Edison schools demonstrated the promise inherent in Edison's model. These schools made productive use of system-level assistance and responded positively to Edison's accountability mechanisms. They were schools with strong instructional leadership, motivated teachers, effective use of achievement data, high-fidelity implementation of the Edison curricula, and high levels of social capital. Yet the realization of this ideal was not universal across Edison schools. With regard to the accountability mechanisms, we found that not all efforts to establish staffing authority or reduce bureaucratic control succeeded. Furthermore, in some schools, the focus on test scores embedded in Edison's monitoring and rewards strategies sparked some of the same responses to high-stakes testing and accountability systems witnessed nationally (e.g., greater attention to tested elements of the curriculum over non-tested elements, focus on the bubble kids), undermining the effort to provide a broad, "world-class" education. As for the resources and assistance offered to schools, teacher attrition at times diminished the value of investments in professional development, and not all schools took full advantage of technology for instructional purposes. In addition, many schools did not achieve the ideal model of distributed leadership and struggled to find teachers to take on leadership roles, and to find time for those who did to adequately fulfill these responsibilities.

Among the schools we visited, several factors appeared to be important in explaining some of the variation we observed. In particular,

- Strong instructional leadership by the principal was associated with stronger implementation of the curriculum, in both tested and nontested subjects.
- Among the case study schools, strong instructional leadership by principals appeared to be more prevalent in charter schools than in district schools. But charter status did not appear to be directly related to curriculum implementation.
- Local constraints, sometimes resulting from compromises required by local contracts, sometimes undermined the implementation of Edison's preferred professional environment.
- Full implementation of the Edison design took time. Schools in their first year of operation encountered frequent challenges in implementing various elements of the design. Most Edison schools that had implemented the design for several years had successfully addressed the first-year challenges and were implementing the design with greater fidelity.

We now turn to an exploratory analysis examining school-level conditions associated with student achievement trends in Edison schools.

TABLE 2  
Achievement Distribution of Case Study Schools in Year of Visit Relative to Total  
Distribution of Edison Schools

	Math	Reading
Lowest quartile	3	2
Middle quartiles	5	7
Highest quartile	6	5
<i>n</i>	14	14

### IMPLEMENTATION AND ACHIEVEMENT IN CASE STUDY SCHOOLS

As noted earlier, the Edison elementary schools that we examined as case studies were somewhat higher performing than Edison averages. Nevertheless, they represented a wide range of performance, including all quartiles of the Edison achievement distribution, as represented by their schoolwide proficiency gains (converted to rank-based *z* scores to permit comparability across sites using different state tests, and using the end of the 1st year of Edison operation as a baseline from which to measure gains) for the operation year in which we visited them (as shown in Table 2).<sup>11</sup> This range of performance provides a useful opportunity to examine school-level factors that might be related to achievement, even if only for suggestive purposes, given the small size and lack of representativeness of the sample.

First of all, we examined the relationship between our ratings of curriculum implementation and the school's achievement gain score in the year of the visit. We found that schools that did better implementing the Edison curriculum in reading and math also posted larger gains in those subjects, on the order of 0.3 to 0.5 standard deviations.<sup>12</sup> Given the small sample sizes involved, the differences were not statistically significant.

Of interest, however, in the case study schools, reading scores were also predicted by math implementation, and both reading and math scores were predicted by an index of the implementation of nontested aspects of the curriculum, including science, social studies, specials, and core values. Effect sizes for the relationship between implementation of nontested curriculum and math and reading test scores were on the order of half a standard deviation, which is at least moderate in size,

<sup>11</sup>The sample size for these analyses is less than the total number of case study schools because complete achievement data were not available for all case study schools.

<sup>12</sup>Similarly, Zhang, Shkolnik, and Fashola (2005) found that schools that had been implementing a comprehensive reform model for three to five years and that were rated as strong implementers achieved larger test-score gains than schools of similar vintage that were judged to be low implementing.

TABLE 3  
Mean Achievement Z Scores by Principal Instructional Leadership, Case Study Schools

	Strong Instructional Leaders		Others	
	Z	n	Z	n
Reading	0.44	6	-0.23	4
Math	0.70	5	0.09	4

and substantial by the standards of education research.<sup>13</sup> To be sure, with simple cross-sectional correlations such as these, we cannot conclude that the relationship is causal. The correlations among the different subjects may simply result from the fact that high-performing schools do many things better than low-performing schools. Nevertheless, these results suggest the intriguing possibility that Edison schools may do better in reading and math achievement if they implement the full Edison curriculum in all of its breadth. At minimum, the results suggest that schools do not need to narrow the curriculum to promote strong achievement in math and reading.

The quality of the principal’s instructional leadership appeared to be strongly related to achievement in both reading (where schools of strong principals scored higher by about 0.7 standard deviation) and math (where schools of strong principals scored higher by about 0.6 standard deviation), as indicated in Table 3.<sup>14</sup> Again, this is a result that might be expected (in Edison schools and non-Edison schools alike), and it is difficult to make a causal attribution. Still, the apparent magnitude of the effect is impressive, suggesting that Edison may be right to put substantial effort into identifying, recruiting, and training principals to be effective instructional leaders.

The implementation of the Edison professional environment—including the use of houses, the availability of planning time, and the prevalence of site-based professional development—was also related positively to achievement in the case study schools, with a correlation of about 0.5 in both reading and math. Schools that followed Edison’s design for school organization were seeing greater student

<sup>13</sup>Note that this effect size cannot be directly compared to the achievement Z-score scale, which is standardized relative to a different distribution.

<sup>14</sup>We examined the relationship between instructional leadership and achievement both for the year of the visit and across all operation years (controlling for Edison-wide operation year trends), on the rationale that principal’s instructional leadership might affect both the current level of the school’s achievement and its deviation from general Edison trends in all operation years. Apparent effects on overall trends controlling for operation year are comparable to apparent effects in the operation year of the visit. Sample sizes in Table 3 are somewhat smaller than in other case study analyses because we lack instructional leadership ratings for a few principals (as well as lacking achievement results for some schools).



achievement gains. Again, although we cannot determine that the relationship is causal, the finding provides encouraging support for the importance of the professional environment as an integral part of Edison's school improvement strategy.

We also examined the relationships between two structural or contextual characteristics of the Edison case study schools and student-achievement effects. Because these are characteristics that tend not to change over time, we would not necessarily expect to observe an effect on achievement in the particular year of our site visit, but we might expect to observe an effect on achievement across all operation years, controlling for Edison-wide operation year trends. The first contextual variable of interest is the extent to which Edison schools operate under local contractual constraints. Edison schools that operated with more local constraints on the implementation of the Edison model had slightly worse achievement outcomes in reading trends (about 0.3 standard deviation lower, on average) and in math trends (about 0.2 standard deviation lower, on average). The second variable of interest is the principal's authority over staffing. Schools in which the principal had full authority to hire and fire teachers had slightly better achievement trends in reading (0.4 standard deviation) and in math (0.1 standard deviation). All of these differences are small (and short of statistical significance), so they should be viewed only as suggestive, but all are consistent with the view that Edison achieved better results in schools where it could fully implement its design.

## IMPLICATIONS AND CONCLUSIONS

Our findings suggest several implications for external organizations choosing similar approaches to improving public schools as well as districts seeking to work with EMOs and other partners. First, there is good evidence from the Edison case study schools that principals' instructional leadership is directly related not only to effective implementation of Edison curricula but also to student achievement. Efforts to identify, recruit, and train effective instructional leaders in the principalship appear to be critical to any organization's efforts to improve the quality of teaching and learning. Other research similarly finds that principals' instructional leadership is related to the likelihood of school change and student learning (see, e.g., Leithwood, Louis, Anderson, & Wahlstrom, 2004; Waters, Marzano, & McNulty, 2003).

Second, our research indicates that multiple accountability systems may influence and in some cases impede school-level improvement efforts. In all cases, Edison's accountability system was not implemented in a vacuum but instead layered on existing state and/or local accountability systems. Edison's accountability system created additional incentives to raise test scores but included other elements of accountability as well. External organizations such as Edison and district staff

need to understand the pressures facing schools and the extent to which the goals imposed on schools by the state, the district, and Edison or other external partners are compatible with one another. For example, we heard from some Edison staff that the professional development workshops the district required them to attend often emphasized topics and goals that conflicted with what Edison expected them to do. Even when it came to achievement goals, the district's expectations could deviate from Edison's, particularly with respect to the relative emphasis placed on status versus growth measures. Administrators need to examine whether undesirable incentives are created by these multiple accountability systems and whether these incentives work to undermine improvement efforts (scenarios we did not widely observe but are clearly possible if not monitored properly). If undesirable incentives are identified, districts, for example, can work to address them through training or through modification of their own approaches for motivating and rewarding school staff.

Third, our findings, like those of other studies (Darling-Hammond, 1988, 1995, 1997; G. A. Hess, 1995;Sizer, 1992) indicate that significant change takes time. Districts partnering with external organizations cannot expect instant improvement. It is important that everyone involved in the decision to bring in an EMO or intermediary partner understand that desired results might not materialize for a few years and commit to sustained partnerships over several years. Our data also indicate that support and oversight are critical during the first year of becoming an Edison school. Although Edison provided extensive professional development, our interview participants told us they would benefit from additional, ongoing support throughout the year. The challenges of the first year were apparent in start-up schools (typically charter schools) and conversion schools (typically district contract schools) alike—ranging from difficulty implementing new curriculum to using technology to operationalizing facilities to filling leadership positions. One promising approach might be to build and strengthen interactions between staff at new schools and staff at existing, successful schools by facilitating mentoring relationships, arranging for instructional leaders in new schools to spend time in existing schools prior to the first year, and encouraging a small, select group of educators from existing schools to transfer to new schools.

Finally, it is important to interpret the findings of our study of Edison Schools in the context of other efforts to improve teaching and learning, particularly in schools facing long-term problems. These include the implementation of comprehensive school reform models (some of which are discussed in other articles in this issue), partnerships with other education management organizations and intermediary organizations, and district and state reconstitution policies. Our broader analysis (see Gill et al., 2005) examined a set of matched comparison schools, but we lack information on what kinds of reform efforts were being undertaken in those schools. There is little information yet available on whether any of these alternate approaches leads to short-term or long-term gains, or how the period required for

Edison to surpass other schools' performance compares with the time trajectories of other approaches. The results provided by this study should serve as a catalyst for additional, comparative research on Edison and other approaches to school improvement.

## REFERENCES

- Berends, M., Kirby, S. N., Naftel, S., & McKelvey, C. (2001). *Implementation and performance in New American Schools: Three years into scale-up* (MR-1145-EDU). Santa Monica, CA: RAND. Retrieved from <http://www.rand.org/publications/MR/MR1145/>
- Bodilly, S. (1998). *Lessons from New American Schools' scale-up phase* (MR-942-NAS). Santa Monica, CA: RAND. Retrieved from <http://www.rand.org/publications/MR/MR942/>
- Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal*, 42, 231–268.
- Chubb, J. E. (2004). The first few years of Edison Schools: Ten lessons in getting to scale. In T. K. Glennan, S. J. Bodilly, J. Galegher, & K. Kerr (Eds.), *Expanding the reach of reform: Perspectives from leaders in the scale-up of educational interventions* (MG-248-FF, pp. 487–516). Santa Monica, CA: RAND.
- Chubb, J. E., & Moe, T. (1990). *Politics, markets, and America's schools*. Washington, DC: Brookings Institution.
- Cookson, P. W., Embree, K., & Fahey, S. (2000). *The Edison Partnership Schools: An assessment of academic climate and classroom culture*. New York: Teachers College, Columbia University.
- Darling-Hammond, L. (1988). Policy and professionalism. In A. Lieberman (Ed.), *Building a professional culture in schools* (pp. 57–77). New York: Teachers College Press.
- Darling-Hammond, L. (1995). Policy for restructuring. In A. Lieberman (Ed.), *The work of restructuring schools: Building from the ground up* (pp. 157–175). New York: Teachers College Press.
- Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work*. San Francisco, CA: Jossey-Bass.
- Edison Schools. (2002a). Assessment and accountability. *Edison Pages*.
- Edison Schools. (2002b). Student management. *Edison Pages*.
- Edison Schools. (2004). *Sixth annual report on school performance, 2002–2003*. New York: Author.
- Edison Schools. (n.d.). *The ten fundamentals behind Edison's School design*. New York: Author. Retrieved June 15, 2005, from <http://www.edisonschools.com/design/designdefault/d0.html>
- Gill, B., Hamilton, L., Lockwood, J.R., Marsh, J., Zimmer, R., Hill, D., et al. (2005). *Inspiration, perspiration, and time: Operations and achievement in Edison Schools* (MG-351-EDU). Santa Monica, CA: RAND. Retrieved August 21, 2007, at [http://www.rand.org/pubs/monographs/2005/RAND\\_MG351.pdf](http://www.rand.org/pubs/monographs/2005/RAND_MG351.pdf)
- Gomez, J. J., & Shay, S. A. (2000). *Evaluation of the Edison Project School: Third interim report: 1998–99 school year*. Miami, FL: Miami-Dade School District.
- Government Accountability Office. (2002). *Insufficient research to determine effectiveness of selected private education companies* (GAO-03-11). Washington, DC: United States General Accounting Office. Retrieved August 21, 2007, from <http://www.gao.gov/cgi-bin/getrpt?GAO-03-11>
- Hamilton, L. S., Stecher, B. M., Marsh, J., McCombs, J. S., Robyn, A., Russell, J., et al. (2007). *Implementing standards-based accountability under No Child Left Behind: Responses of superintendents, principals, and teachers in three states* (MG-589-NSF). Santa Monica, CA: RAND. Retrieved August 21, 2007, from [http://www.rand.org/pubs/monographs/2007/RAND\\_MG589.pdf](http://www.rand.org/pubs/monographs/2007/RAND_MG589.pdf)

- Hentschke, G., Oschman, S., & Snell, L. (2003). *Trends and best practices for education management organizations*. San Francisco, CA: WestEd. Retrieved August 21, 2007, from [http://www.wested.org/online\\_pubs/PP-03-02.pdf](http://www.wested.org/online_pubs/PP-03-02.pdf)
- Hess, F. M. (1999). *Spinning wheels: The politics of urban school reform*. Washington, DC: Brookings Institution Press.
- Hess, G. A. (1995). *Restructuring urban schools: A Chicago perspective*. New York: Teachers College Press.
- Hill, P. T., Pierce, L. C., & Guthrie, J. W. (1997). *Reinventing public education: How contracting can transform America's schools*. Chicago, IL: University of Chicago Press.
- Kirby, S. N., Berends, M., & Naftel, S. (2001). *Implementation in a longitudinal sample of New American Schools: Four years into scale-up*. Santa Monica, CA: RAND.
- Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *Review of research: How leadership influences student learning*. Minneapolis, MN: Center for Applied Research and Educational Improvement, University of Minnesota.
- Miron, G., & Applegate, B. (2000). *An evaluation of student achievement in Edison Schools opened in 1995 and 1996*. Kalamazoo, MI: The Evaluation Center of Western Michigan University. Retrieved August 21, 2007, from [http://www.wmich.edu/evalctr/edison/wmu.edison\\_rpt.pdf](http://www.wmich.edu/evalctr/edison/wmu.edison_rpt.pdf)
- Molnar, A., Garcia, D. R., Bartlett, M., & O'Neill, A. (2006, May). *Profiles of for-profit education management organizations: Eighth annual report, 2005–06*. Tempe, AZ: Commercialism in Education Research Unit, Arizona State University.
- Pedulla, J. J., Abrams, L. M., Madaus, G. F., Russell, M. K., Ramos, M. A., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Chestnut Hill, MA: National Board on Educational Testing and Public Policy.
- Rhim, L. M. (April 2002). *School privatization by way of a comprehensive management contract: a single case study of the extent to which privatization theory transfers to practice in a public charter school*. Paper presented at the American Educational Research Association, New Orleans.
- Sizer, T. R. (1992). *Horace's school: Redesigning the American high school*. Boston, MA: Houghton Mifflin.
- Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced leadership: What 30 years of research tells us about the effect of leadership on student achievement*. Aurora, CO: Mid-continent Research for Education and Learning.
- Zhang, Y., Shkolnik, J., & Fashola, O. (2005). *Evaluating the implementation of Comprehensive School Reform and its impact on growth in student achievement*. Washington, DC: American Institutes for Research. Retrieved August 21, 2007, from <http://www.air.org/news/documents/AERA2005Implementation%20of%20CSR-Impact%20on%20Student%20Achievement.pdf>



# Melding Internal and External Support for School Improvement: How the District Role Changes When Working Closely With External Instructional Support Providers

Jonathan Supovitz

*University of Pennsylvania Consortium for Policy Research in Education*

To support instruction, school districts must provide a wide array of assistance to schools. Broadly speaking, districts play the roles of *authority* in holding schools accountable for their activities and performance, *support* in assisting school faculties to build their capacity to better instruct students, and *brokerage* between schools and outside providers of service and materials. The roles of authority, support, and brokerage typically contend with each other, producing a set of perennial tensions for district leaders. This article examines the influence on these three roles of external support providers working in close partnership with districts on instructional improvement efforts. First, the article reviews the literature on district/provider partnerships for examples of role adjustment. Second, using a case study of a deep partnership between a district and an external provider, this article empirically examines the influence of a district/provider partnership on the balance of district roles. The findings illustrate how the traditional district roles of authority, support, and brokerage are adjusted by partnerships with external providers.

Productive partnerships between school districts and external education service providers are underutilized resources for instructional improvement in education today. To support instruction, districts must provide a wide array of assistance to schools. The breadth of these demands requires that districts themselves seek reinforcement. The involvement of external providers can provide needed support but changes the dynamics of district's relationships with schools. Broadly speaking,

districts play the roles of *authority* in holding schools accountable for their activities and performance, *support* in assisting school faculties to build their capacity to better instruct students, and *brokerage* between schools and outside providers of service and materials. The roles of authority, support, and brokerage contend with each other, producing a set of perennial tensions for district leaders. Using a case study of a deep partnership between a district and an external provider, this article examines the influence of district-provider partnership on district roles of authority, support, and brokerage.

## DISTRICT SUPPORT FOR INSTRUCTIONAL IMPROVEMENT IN SCHOOLS

Districts provide—to a more or less effective extent—an array of support functions for schools. Cuban (1988) studied the roles of district leaders and identified three sets of responsibilities, which he labeled administrative chief, negotiator-statesman, and instructional supervisor. Administrative chiefs were foremost committed to directing organizations that were dedicated to achieving the highest levels of productivity and efficiency. Negotiator-statesmen considered community relations and the political dimensions of their jobs to be especially central. Instructional supervisors emphasized themselves principally as “teachers of teachers” and therefore viewed classroom support as their primary function (Cuban, 1988, p. 112). Extrapolating these roles out to the organization, district support functions can be similarly grouped into the managerial, political, and instructional. Cuban argued that, historically, the political and managerial functions overwhelmed leaders’ attention to instruction. Thus instructional attention is constantly at risk in the tug of war for district leaders’ attention.

Within the instructional realm, districts also supply different types of support for schools. Supovitz (2006) identified seven instructional support functions that districts can supply to help schools in their efforts to enhance the quality of teaching and learning. These seven instructional support functions are briefly described next.

1. *Coordinator of curriculum and instructional materials.* Most of the established sets of textbooks and materials that dominate education come from a set of national producers and exhibit remarkably little variability (Goodlad, 1999). However, the curricula that children experience are made up of a broader set of influences (Tyler, 1988). These include such things as curriculum guides; instructional materials; the scope and sequence of lessons within units; and the array of supplemental materials that include tasks and kits, lesson plans, pacing guides, and assignments (Ball & Cohen, 1996; Goodlad, 1999). Districts have a long history of developing and/or supplementing the materials that

teachers use in their classrooms and defining the sequence by which those materials are used (Tyack & Cuban, 1995). Text adoptions are the primary routine in most districts for updating the curriculum (Carus, 1990). Districts also play a major role in providing curricular guidance and coherence (Massell, 2000).

2. *Professional development provider.* Building the capacity of teachers and school leaders to deliver and support powerful instruction to students is a major need of school faculties, and districts play a critical role in both setting the context for and providing professional development (Knapp, Zucker, Adelman, & St. John, 1991; Spillane & Thompson, 1997). Districts are the predominant deliverers of professional development for teachers (Firestone & Hirsh, 2005) and typically spend anywhere from 3 to 7% of their budgets on school faculty professional development (Miles, Odden, Fermanich, & Archibald, 2005). Research on district professional development has found much variability in its coherence (Spillane, 1996). Summarizing district management of professional development, Desimone, Porter, Birman, Garet, and Yoon (2002) found that higher quality professional development was associated with such district strategies as aligning professional development to standards and assessments, continuous improvement efforts, and teacher involvement in professional development planning.
3. *Monitor of program implementation.* Program evaluation is a critical technique to ensure the efficacy of intended interventions (Rossi, Freeman, & Lipsey, 2003). Districts often conduct mandatory evaluations as required by government and funding agencies (King, 2002). Beyond required evaluations, when school or district leaders identify a practice that they believe to be effective and invest in materials and resources associated with that practice, they should be interested in understanding the extent to which those practices are being adopted in schools and classrooms and the extent to which programs and practices are contributing to the learning of students. A national survey conducted in the 1980s (Banks & Williams, 1981) found evaluations to be a common, but weak, function in medium- and large-size school districts. With the advent of the standards and accountability movements, attention to building evaluation capacity building in school districts is growing (King, 2002).
4. *Organizer and deliverer of student performance results and other data to inform instructional and strategic decision making.* The collection and disaggregation of student performance data is becoming more and more prevalent in school districts (Herman & Haertel, 2005). Teachers and school leaders are increasingly being asked to make decisions based on a range of data (Datnow, Park, & Wohlstetter, 2007; Earl & Katz, 2002). The imperfect alignment between assessment for accountability purposes and for formative feedback to teachers is giving way to a host of formative and interim benchmark assessments that provide finer grained information to teachers. Schools generally lack the

resources and technical expertise to coordinate the increasing amounts of data available to them (Supovitz & Klein, 2003). The demands associated with the emphasis on data-driven decision making raise a host of issues for districts and schools. First, someone must organize the data so that decision makers have access to what they need in a form that is useful. Many districts are working with external providers to provide data management and warehousing functions (Wayman, Stringfield, & Millard, 2004). Second, increased technical capacity requires an increase in the human skills necessary to turn data into actionable knowledge (Boudett, City, & Murnane, 2005; Petrides & Guiney, 2002).

5. *Searcher for ideas, high-quality materials, programs, and practices to bring into the system.* Organizational improvement is built upon the infusion of new ideas and better ways of doing things (Fullan, 2005; Senge, 1990). School districts are continually searching for ways to support and enhance their instructional programs. Social science researchers have long studied how organizations identify their problems and search for solutions (Cyert & March, 1963; Simon, 1979). Researchers have identified key elements of the search process, including the decision situations, participants, and problems that are identified (Cohen, March, & Olsen, 1972). In the literature, the search for innovation is associated with improved organizational performance (Ahuja, 2000; Stuart, Hoang, & Hybels, 1999). Within education, the search situation is made more complex by many of the inherent difficulties in the education industry, namely, that good information is difficult to come by and design and measurement problems obfuscate the true merit of educational interventions (Mosteller & Boruch, 2002). Although there are several efforts underway to consolidate stable information about program quality (i.e., What Works Clearinghouse; Comprehensive School Reform Quality Center), this is an endemic problem in education. Effective searching requires the ability to distinguish between differing levels of program quality, having a broad perspective on the industry's landscape, and having the time and resources to scour that landscape (Daft & Weick, 1984; Levinthal & March, 1981). School leaders lack the time and expertise to effectively search for innovations and districts generally play this role (Gross, Kirst, Holland, & Luschei, 2005).
6. *Facilitator of networks between schools as a mechanism for spreading and sharing knowledge.* The task of education is very consuming for school leaders and faculties, and schools can become isolated in their efforts. For this reason, schools can benefit from structured opportunities that allow teachers and leaders to share with and learn from the experiences of their peers (DuFour & Eaker, 1998). Fullan (2005) argues that network structures are a critical element to school improvement. "We can't change the system without lateral (cross-school and cross-district) sharing and capacity development," he contended



(p. 66). Thomas (2004) documented the growth of leadership networks within districts as a strategy for improving school leadership. Elmore and Birney (1997) identified district-sponsored peer networks as an important contributor to professional learning and instructional improvement in New York City's Community District 2.

7. *Coherer of programs and resources.* Finally, but not least, districts coordinate the range of activities, resources, and policies that engage schools. One of the chief findings that came out of decades of programmatic research on educational reforms is that individual reforms are likely to be ineffective if they are implemented in isolation amidst other incompatible efforts (Fuhrman, 1993; Smith & O'Day, 1991). Districts must consider the "fit" between programs and policies to encourage compatibility and synergy and assure philosophical alignment (Fuhrman & Massell, 1992; Kahle, 1997). Therefore, a crucial role of an effective education support organization is to orchestrate amongst particular programs and to provide coherence across them (Floden, Goertz, & O'Day, 1995).

The broad range of these instructional support responsibilities makes the challenge of improving teaching and learning daunting. Effectively providing leadership in these instructional support areas, particularly when combined with their managerial and political responsibilities, presents a challenge to school districts. To better meet these responsibilities, districts themselves seek a range of support and expertise.

One thing often overlooked in examining district support for instructional improvement is the important role of external assistance in this process. In fact, many of the instructional support tasks just described require districts to manage, coordinate, and integrate products and services developed and supported by external providers. For example, much curriculum and related materials, as well as their associated professional development, are externally developed by textbook publishers and other entities. In addition, other programs, whether they be technology additions, data use systems, discipline or safety programs, or dropout prevention interventions, to name just a few, are often externally developed and must be accommodated into district routines and practices.

Beyond programs, external support often takes on other forms. For example, intermediary organizations play an important role in instructional improvement within the district context (Cervone & McDonald, 1999; Honig, 2004). In other cases, subject matter networks, such as the National Writing Project, work with districts and schools to improve literacy (Lieberman & Wood, 2003). Thus a major component of the instructional function of districts arises from efforts to broker between school needs and external products and services.

## THE ROLE OF EXTERNAL SUPPORT PROVIDERS IN THE LITERATURE ON DISTRICT INSTRUCTIONAL IMPROVEMENT

A review of the research on district improvement with an eye toward the role of external assistance reveals an interesting pattern. Most research either focuses on the district role in instructional improvement in schools, barely mentioning the contribution of external assistance in these efforts, or emphasizes the improvement efforts through the lens of the external provider, minimizing the district role. Only in a few current cases are we starting to see research that explores the partnerships and relationships between districts and external support providers.

Most recent research on districts focuses on district strategies and their effects and makes little mention of the role of external providers. Two notable examples serve to illustrate the thin treatment of providers in district improvement efforts. A highly publicized report from The Manpower Demonstration Research Corporation (Snipes, Doolittle, & Herlihy, 2002) presented case studies of four urban systems that were improving student achievement. The researchers selected the districts based on trends of improvement in reading and mathematics from 1995 to 2001. Their report highlighted the need for a prolonged period of political and organizational stability and consensus on educational reform strategies. They found that the improving districts shared several things in common including a focus on student achievement and specific achievement goals, aligned curricula with state standards, and translated standards into instructional practices; a well-specified system for holding district leaders and building staff responsible for producing results; a focus on the lowest performing schools; districtwide curricula and instructional approaches; clearly defined central office roles; and a commitment to data-driven decision making and instruction. Although the authors noted curricular, coaching, data use, or other externally developed programs and materials—and the support surrounding them—in their study, they did not focus on the relationships between the providers of these external resources and the district's efforts. Thus, the range of resources that the authors described that the successful districts employed were mostly transparent in this study.

As a second example of how the roles of external providers were only discussed in passing in district improvement efforts, Togneri and Anderson (2003) examined the traits of five high-poverty districts that were improving the achievement of their students. The authors found that the districts had “a strikingly similar set of strategies to improve instruction” (p. 4). These included the courage to acknowledge poor performance and the will to seek solutions; a vision that focused on student learning and guided instructional improvement; a systemwide approach to improving instruction, including curricula and instructional supports; data-based decision making; new approaches to professional development; redefined leadership roles; and commitment to sustaining reform over the long haul. Although

the authors mentioned some work of external providers in passing, including districtwide curriculum, data use programs, external funders, and university partners, they did not explicitly explore the role of providers in district improvement efforts.

There are two possible conclusions to reach from the small mention of providers in these and other pieces on district improvement. First, it is possible that many districts are largely operating without external assistance. Alternatively, it is possible that the provider role was present but not much discussed. I tend to think it is more the latter for two reasons. First, because ever since the late 1980s and early 1990s, district central offices have been dramatically downsized (Leithwood, 1995; Mac Iver & Farley, 2003; Payzant & Gardner, 1994). So who is doing the development of instructional materials and coordinating professional development? Second, the market of external support providers is tremendous—one only need ask district administrators and school principals about the amount of solicitations they get from vendors or look at the advertisements in trade magazines and journals to recognize that this market is robust.

A series of efforts to support school improvement from outside the system in the 1990s expanded the role and raised the visibility of external support organizations in American education. The 1993 \$500 million Annenberg Challenge grant gave rise to a series of intermediary assistance organizations in major U.S. cities, including Boston, Chicago, Houston, Philadelphia, Los Angeles, and San Francisco (Kronley & Handley, 2003; Schön & McDonald, 1998). The Comprehensive School Reform movement in the 1990s further spurred the development of external organizations providing instructional support to schools and districts (Bodilly, 1998; Borman, Hewes, Overman, & Brown, 2003; Desimone, 2002).

These developments are giving rise to deeper investigations of the relationships between districts and external providers in the literature on school reform. Researchers are increasingly acknowledging the role of external support providers in districts' instructional improvement efforts and probing the relationships between districts and providers as they explore the processes that districts go through in their efforts to improve instructional quality. Richer descriptions of how district leaders and providers interact and work together to support schools are emerging. Here I describe three examples to illustrate the tenor of the relationships represented in the literature.

RAND researchers have studied the partnerships between the Institute for Learning (IFL) and three urban districts (Marsh et al., 2005) from 2002 to 2004.<sup>1</sup> The IFL is a nonprofit organization coordinated by the Learning Research and Development Center at the University of Pittsburgh. In the study districts, the IFL work focused on the development of instructional leadership, school-based

---

<sup>1</sup> See also Honig and Ikemoto (2008/this issue).



coaching, curriculum specification, and data use. The researchers employed a comparative case study design that featured two years of fieldwork, focus groups with teachers, surveys of principals and teachers, analysis of IFL documents, and analysis of student achievement data. The researchers found that the IFL affected the organizational culture, norms, and beliefs about instruction and helped develop the knowledge and skills of central office administrators. They discussed the several lessons stemming from their observation of the district-provider relationship. These included the importance of strong relationships at all levels of the organization to enable partnership efforts, the wariness of local faculties of the reputation of external vendors, the importance of provider credibility and tools to build support in schools and at the district level, the influence of the pre-existing reform context on a new partnership relationship, the constraints of the capacity of the provider and its services in relation to larger district needs, and the extent to which the providers offerings align themselves to broader district needs.

The Manpower Demonstration Research Corporation (MDRC) examined a four year partnership between the Institute for Research and Reform in Education's First Things First (FTF) reform model and the school district of Kansas City, Kansas, funded and facilitated by the Kauffman Foundation (Gambone, Klem, Moore, & Summers, 2002; Quint, Bloom, Black, Stephens, & Akey, 2005). The researchers depict FTF as a "theory of change" approach to districtwide school reform in which reforms and districts form a close partnership and agree to both the strategies and sequences that the reforms will take and to the responsibilities of each party. The FTF reform features small learning communities in which students stay over multiple years; a family advocacy system in which staff members meet with students and monitor their academic, social, and emotional progress; and standards-based faculty instructional improvement efforts. FTF's close partnership with Kansas City, a district with 21,000 students in 47 schools, featured a careful phase-in sequence for school feeder patterns, the delivery of intensive professional development, and the reassignment of district-level curriculum specialists into school improvement facilitators to lead the change process in schools. The authors documented several early outcomes in the district, which they believed to be forerunners of improved academic performance, including increases in stakeholder awareness and knowledge of the reform, commitment to implement the reform, and readiness for implementation of critical FTF features. Subsequent MDRC evaluation reports showed substantial effects of the reform on improving a wide range of academic outcomes in Kansas City (Quint et al., 2005). The MDRC authors explored many aspects of the reform efforts. In terms of the Kansas City/FTF partnership, the authors stressed three important points. First was the close and flexible partnership between the district and provider. Second was the district's active provision of both pressure and support for the reform. Third, the authors emphasized the intensive and responsive technical assistance from FTF, who were willing to make adjustments when needed.



A third example illustrates how provider relationships are emerging in the discourse around district improvement. In 2002, Hightower, Knapp, March, and McLaughlin produced an edited volume entitled *School Districts and Instructional Renewal*. The book featured many important topics including district relationships with states, schools and communities; the internal capacity building strategies of District 2 in New York City; San Diego, California; New Haven, Connecticut; and district leadership. The role of external assistance is not well represented in the volume but is mentioned as one lesson in the conclusion to the volume. Therein the authors “underscore the ways in which the district’s capacity to learn, lead, and educate is enhanced by partnerships with external organizations.” The authors continued, “these partnerships not only expand the capacity of the district but also extend the concept of ‘district’ to include a particular community and professional context. Conceived in this way, new avenues arise for districts to develop more consequential approaches to instructional renewal” (p. 196).

#### EXAMPLE OF A CLOSE PARTNERSHIP: DUVAL COUNTY, FLORIDA, AND THE NATIONAL CENTER ON EDUCATION AND THE ECONOMY

In 2006 I completed an extensive longitudinal case study on the reform efforts of the Duval County Public Schools (DCPS) from 1998 to 2004 and the positive impacts on student achievement (Supovitz, 2006). DCPS is the 20th largest urban school district in the nation, educating approximately 130,000 children in 150 schools. During the time of the study, the district was led by John Fryer, a retired U.S. Air Force major general with no formal experience leading schools. One of the unusual aspects of the DCPS reforms during Fryer’s tenure was the district’s extensive partnership with the National Center on Education and the Economy (NCEE) to provide instructional assistance to the district. NCEE is a non-profit education support provider headquartered in Washington, DC.

Led by Marc Tucker and Judy Coddling, NCEE has been one of the nation’s foremost education support providers over the past 20 years. Going back to the 1990s, Tucker and Lauren Resnick led the New Standards Project, one of the early efforts to produce national standards in the major subject areas. In the late 1990s, as part of the New American Schools development program, NCEE developed America’s Choice, one of the major Comprehensive School Reform programs. By 2003, America’s Choice was operating in about 450 schools across the country (Borman et al., 2003).

Over Fryer’s tenure, Duval County became one of the largest sites for America’s Choice, and Fryer developed a close relationship with NCEE’s leaders. In 1999, Duval County began implementing America’s Choice in 14 schools. Subsequently, the district expanded America’s Choice to roughly one third of the district’s

150 schools. The introduction of America's Choice in schools involved direct professional development of principals and two school coaches by NCEE and monitoring of the reform in schools by NCEE-employed regional cluster leaders.

Over time, the partnership between the district and NCEE evolved further. Although the district was implementing America's Choice, it began literacy coach training in all the other schools in the district not implementing America's Choice. In 2004 the district became a pilot site for NCEE's leadership training program, the National Institute for School Leadership. As formal America's Choice adoption ended in the district, DCPS entered into a licensing agreement to use NCEE training materials and paid NCEE to train and certify its district standards coaches to oversee schools' continued implementation of standards-based reform.

At times, the relationship between the district and NCEE was uneasy, as both parties sought to define their appropriate roles in support of the school improvement process. Although Duval County leaders recognized the need for, and value of, NCEE's abundant expertise, they also perennially tried to bring these qualities in-house and gain independence from external reliance. Throughout the evolution of the partnership with NCEE, Duval County leaders repeatedly sought to bring the expertise provided by NCEE into the district's functions to become self-sufficient, only to find that they were better off, for a variety of reasons, to continue to utilize the external expertise. When it came to NCEE's expertise, Duval County leaders were always asking themselves, "Should we buy it or should we make it?"

The case of whether to purchase or develop instructional training materials for teacher professional development is a particular case in point to illustrate the pendulum of Duval County's leaders' thoughts as to what the role of external providers ought to be relative to the district's function. In the space of four years, from 2000 to 2003, Duval County went from a purchaser of NCEE training and development materials to the developer of those materials, then back to the purchaser. Even as they returned to purchasing, they were concerned with how well the materials were suited for their context and sought to customize them.

When Duval County leaders initially adopted America's Choice, they had every intention of using the program as a way of building the district's internal capacity. As Ed Pratt-Dannals, the associate superintendent for curriculum and instruction at the time (and now superintendent) said in 2002, "From the beginning there was an agreement with the Superintendent (Fryer) and Judy Coddling (of NCEE) that the ideal would be that over a three to five year period we'd be creating internal capacity."

Superintendent Fryer, using a pilot's metaphor, expressed the same idea in a December, 2002 interview:

I told [NCEE] from the beginning that I am not interested in a model where I have to stay connected by an umbilical cord forever. I wanted their capabilities for a fast takeoff, rather than a slow climb. I saw what other districts had done. I saw what was

going on in New York's District 2 and I saw that it took them 10 years to build and understand standards and I didn't have 10 years. I wanted to get going.

So, as NCEE was training teachers and school principals in Duval County to implement America's Choice, they were also training the people that Fryer hoped would build Duval County's capacity to take over NCEE's role as the deliverer of high quality instructional materials in the district. By the fall of 2002, although Fryer was disappointed at the slow pace of internal development, he still believed that the district was on track to build its own capacity. As he explained in an interview,

My plan was that . . . by the end of three years we would have our own capacity to continue this work in the rest of the schools. It didn't work out that way. . . . I saw quickly, by the end of the second year, that we weren't going to have a capacity if I didn't create one. . . . So I created a team of the very best people we had in our reform and made them trainers and asked them to develop parallel materials that were not proprietary. . . . And you can put all of that together and have a pretty good . . . reform model, and we did that. So we've moved . . . in developing our own capacity and letting our best people put together programs and it appears to be working quite well.

Throughout 2002 and the first half of 2003, a cadre of Duval County's teachers and school leaders developed a set of materials that were in many ways similar, yet did not infringe upon, NCEE's materials. By many accounts, the quality of the materials and training developed by Duval County curriculum developers was high, as was the price.

It was based not on quality but on economic grounds that Duval County leaders relinquished their efforts to make professional development and curricular materials. As Fryer concluded in the spring of 2003,

I didn't want to pay the internal price of having our people develop materials. We costed out what we did last year. It probably cost a couple of hundred thousand dollars for materials to develop them. And you figure it from their time—it's not worth it when for \$700,000 I have eternal right to [NCEE's]. So you know, I'd rather do it with them.

So through a licensing agreement, Duval County continued to license the rights to use NCEE's materials. But even so, Duval County demanded the right to "Duvalize" NCEE's materials, modifying them to fit the district's context and need.

Despite the eventual decision to outsource instructional materials, professional development around those materials was a different story. Using private funding, the district offered professional development to teachers and principals through the Schultz Center for Teaching and Leadership, a nonprofit independently operated



regional training facility. Working closely with the district, the Schultz center staff (many of whom are district employees) provided a highly scaffolded sequence of teacher training in the content areas that aligned with the district's instructional vision and the NCEE curricular materials.

Beginning in 2002, one of the central capacity building strategies of the district was the placement of a full-time standards coach in each school. Duval County leaders absorbed the idea of school-embedded professional development from the America's Choice design. The coach's job was to work with the school principal to implement the district's standards-based reform vision.

In the summer of 2003, Duval County and NCEE contractually extended their relationship through what they called a "design license agreement." The agreement called for the certification of district standards coaches to support the school standards coaches in their implementation of standards-based reform in Duval County schools and gave Duval County the right to use NCEE's copyrighted training, curriculum materials, and the NCEE's detailed school implementation rubrics. Most of the district standards coaches had deep experience with America's Choice. Like America's Choice cluster leaders, district standards coaches were assigned to work with between six and ten schools to support their implementation of the components of the districts frameworks. The coaches were under the supervision of regional superintendents of the district.

Several things are notable about the district's experience with NCEE around instructional expertise. First, the district expended considerable resources to develop instructional and professional development materials but found they could not do this at equivalent quality for the same cost as NCEE. Thus, the district chose to outsource most of its materials development, even while maintaining the right to customize the materials. Second, the district kept in-house the training associated with the materials development while working with NCEE to oversee and certify the quality of the training.

### TENSIONS INHERENT IN DISTRICT ROLES OF AUTHORITY, SUPPORT, AND BROKERAGE

At the heart of a district's efforts to support instructional improvements is to balance roles of authority, support, and brokerage. As an authority figure with the power to hold schools accountable, districts may tend to preside over them, attempting to control their efforts. In a supporting role, with the responsibility to assist schools to improve, districts must take on a different mantle, encouraging and nurturing the efforts of school faculties. As a broker between external providers and schools, intent on identifying and introducing new ideas into schools and classrooms, districts must play more the part of the matchmaker and chaperone.



Each of these roles requires that districts assume a different relationship with schools and the wider world, and at times these roles are in conflict.

Particularly difficult is the uneasy balance between authority and support. It is often difficult for authority figures to provide support because of the antithetical mindsets between ruling (authority) and serving (support). Further, by taking on responsibility for providing support, an authority becomes partly responsible for performance; thereby putting itself in the awkward position of becoming the *target*, as well as *agent*, of accountability. By distancing itself from the target of its support, an authority may clarify its accountability role, but in doing so reduces the quality of its support. Alternatively, the stronger the bond created through a support relationship, the more responsibility is shared, which makes authority more difficult to maintain.

Further, when districts act as brokers between external providers and schools, they may find themselves in the awkward position of effectively diminishing their authority. To do their accountability work, districts must gain the requisite levels of expertise in new reforms to effectively monitor implementation. The credibility of any authority resides in both its formal position and its knowledge and expertise. District expertise may be undermined by relying on an external partner to deliver reforms to schools. If districts are to monitor implementation, support schools for improvement, and hold schools accountable for their progress, they must have the expertise necessary to distinguish between different levels of implementation. In the traditional model, where districts control the intervention and training, they hold the expertise and the schools are the recipients of that expertise. But in situations in which expertise comes from external providers directly to schools, the district is in a potentially un-credible position vis-à-vis schools in that they hold responsibility for implementation yet lack the widespread expertise to distinguish between levels of implementation.

The case of Duval County illustrates how the authority and support relationships in the district were fundamentally changed by the introduction of external reform expertise. The district's regional superintendents and directors had been traditionally responsible for both supporting and monitoring the schools in their region to improve and implement the districts reforms. However, as America's Choice and NCEE's particular instructional approaches were introduced, the regional superintendents and directors found themselves unqualified to oversee schools' implementation of the reforms. The regional administrators had mostly spent their careers in the district and had always been involved in developing the district's instructional reforms, but in this case they lacked the expertise to play either the support or authority roles. The school standards coaches and principals that were trained by America's Choice found themselves with greater knowledge and expertise in the particulars of the reforms than did the district supervisors who were supposed to assist with and monitor school implementation. This left district administrators in the awkward position of being the formal

supervisor of schools but lacking both the knowledge and credibility to play that role.

In essence, the introduction of the external reform changed the expertise equation, and therefore the power dynamic, in the district by shifting the center of knowledge in the system from the district to NCEE. Whereas the expertise for major instructional interventions had traditionally flowed from the central office down to the schools, expertise was now entering into the district from the external provider. What was unusual about this situation was not the direct training of school personnel by external providers, which had been noted by other researchers (Datnow, Hubbard, & Mehan, 2002; Glennan, Bodilly, Galegher, & Kerr, 2004), but by both its magnitude and the way that the external provider and district enfolded their authority and support together.

As a consequence of the changed power dynamics, the situation created strong demand on the part of district administrators to learn about the reform in order to recapture the basis of their authority. Informally and at their own initiative, the regional superintendents and directors sought out training about America's Choice to increase their knowledge and allow them to execute their support and monitoring roles. They began visiting America's Choice schools specifically as professional development to learn about the reforms. These visits evolved into a formal implementation monitoring system that allowed them to recapture their monitoring and accountability function (Supovitz, 2007; Supovitz & Weathers, 2004). Because the reforms persisted in the district for more than eight years, the expertise of the reforms eventually worked its way up into the system through promotion as well. Over time, successful principals of America's Choice schools and America's Choice coaches moved up into district leadership roles as a consequence of their increased knowledge and expertise in the central district reforms.

The district's vacillation about whether to outsource or bring in-house the development of instructional and professional development materials represents another shift in the balance between authority, support, and brokerage. The indecision on the district's part reflects a conflict between the general desire to remain self sufficient while acknowledging the functional reality of the best division of labor. In this case, Duval County found that they could produce quality instructional materials but not as cost efficiently as could NCEE. The fact that the district first tried to produce comparable materials itself, then eventually decided to purchase NCEE's materials, and then afterward persisted in demanding the right to adjust the materials to their own context reflects the district's desire to retain authority and control over the instructional function.

On the other side of the support equation, the district worked out a new equilibrium with NCEE to retain the provision of professional development to teachers and schools within the district. The district largely conducted professional development in-house, using NCEE as a quality control mechanism and a conduit for refined instructional ideas. Most professional development was carried out

through the Schultz Center for Teaching and Leadership. (The Schultz Center itself can even be thought of as another external provider with whom the district was engaging in a partnership.) Thus, relative to NCEE, the district retained the support and monitoring functions around training in-house. However, the district standards coaches, who were NCEE trained and certified, allowed the district to access the continually growing instructional expertise of NCEE. Thus, for the training function, the district played both the roles of authority and support, and NCEE played largely an authority role.

Discussion of the decisions of organizations as to whether to use external services traditionally swings across the fulcrum of retaining services in-house or outsourcing them (Bhagwati, Panagariya, & Srinivasan, 2004; Heshmati, 2003). Such a stark depiction of the choice is probably not the best way to describe a productive relationship between a district and an external provider. Implicit in the concept of outsourcing is that the one initiating the outsourcing (in this case the district) steps back and lets the vender take over the task. But if the thing that is outsourced is a service, as it largely is in the case of professional development, then the outsourcer needs to remain intimately involved because of the ongoing nature of the transaction and the need for sustained support that is integrated with other resources. The introduction of instructional innovation into schools requires ongoing support that must be provided in concert between an external provider and a local entity. Therefore, the act of brokerage may instill a false sense in districts that they need not play a support role. In cases where they do seek to provide support, there may be problems of clarity between their support efforts and those of the external provider. Districts cannot step aside, because they would be abrogating an important component of their responsibility to schools. However, they cannot fully supplant the external provider either, because although they may be able to support the introduced set of practices, they are not in the position to commit the resources to research and development that gave the external provider a competitive advantage in the first place.

## TOWARDS MORE SEAMLESS MELDING OF INTERNAL AND EXTERNAL INSTRUCTIONAL SUPPORT PROVIDERS

This article began with a systematic breakdown of the array of instructional support functions traditionally provided by districts—to a greater or lesser extent—in support of school improvement. Successfully providing this array of functions represents a daunting task for any single education support provider. To more effectively provide these support functions, districts are increasingly entering into sophisticated relationships with external partners.

Underlying these functions is a set of district stances around authority, support, and brokerage that revolve around legitimacy, power, expertise, and trust. The



source of district legitimacy differs depending on whether it is playing the role of authority, support, or brokerage. As an authority, district legitimacy comes from its formal position of power. As a support provider, district legitimacy arises from the expertise and usefulness it provides. As a broker, district legitimacy relies on identifying and bringing in programs and services that are perceived as useful and productive. These different roles also have distinctly different effects on the trust relationships between schools and districts, as hierarchical relationships based on authority tend to reduce trust, whereas support relationships tend to encourage trust. Playing a support role requires that districts develop a high level of expertise to provide meaningful assistance to schools, whereas authority and brokerage require far less.

The advent of more sophisticated partnerships in support of instructional improvement forces districts to reconsider their roles both with schools and external providers and adjust the traditional lines of authority and support within the district context. The examination of the partnership between Duval County and NCEE in this study illustrates several of the issues that arise when districts develop more sophisticated and longer term relationships with external support organizations. The story of the evolution of the roles of Duval County and NCEE around the development of curriculum materials and the provision of professional development reveals some of the adjustments that district administrators undergo as their traditional roles of authority and support change. The shift in responsibilities of providing support to schools and monitoring implementation, as well as the negotiation of these arrangements with external partners, forces district leaders to reconsider their traditional roles of support and authority relative to schools.

As this case study suggests, *partnership* is the operative word in the delivery of instructional services in today's environment. District leaders cannot simply step aside and let an external provider work with schools to support school reform. Rather, they must build an infrastructure to support implementation from different angles than does the provider. These new relationships raise a series of questions. How does provider support fit into the existing structure of the district? How do lines of authority change as expertise comes laterally into the system? How do new initiatives fit into the prevailing program monitoring and accountability structures within the district? How are leaders at different levels of the organization trained to both understand and support the program? How do districts broker these new relationships with external providers? The increasingly prevalent and sophisticated relationships between districts and external providers make these questions important for future investigation.

We might even consider the increasingly sophisticated partnerships that are becoming more commonplace today between districts and external providers as a new model for thinking about how internal and external support providers can work together to provide stronger support for school improvement. By melding their support and comparative advantages together, district and provider partnerships



may be the best option for most effectively supporting instructional improvement in schools.

## REFERENCES

- Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly*, 45, 425–455.
- Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is: Or might be: The role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*, 25(9), 6–8, 14.
- Banks, A., & Williams, R. C. (Eds.). (1981). *Evaluation in school districts: organizational perspectives*. Los Angeles: Center for the Study of Evaluation.
- Bhagwati, J., Panagariya, A., & Srinivasan, T. N. (2004). The muddles over outsourcing. *The Journal of Economic Perspectives*, 18(4), 93–114.
- Bodilly, S. J. (1998). *Lessons from New American Schools' scale-up phase*. Santa Monica, CA: RAND Corporation.
- Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, 73(2), 125–230.
- Boudett, K. P., City, E. A., & Murnane, R. J. (2005). *Data wise*. Cambridge, MA: Harvard Education Press.
- Carus, M. B. (1990). The small publisher in a national market. In K. J. Rehaag, D. Elliott, & A. Woodward (Eds.), *Textbooks and schooling in the United States: Part I. Eighty-ninth yearbook of the National Society for the Study of Education* (Issue 1). Chicago: University of Chicago Press.
- Cervone, B. T., & McDonald, J. (1999). *Preliminary reflections on the Annenberg Challenge: A working paper drawn from its first projects*. Providence, RI: Annenberg Institute for School Reform.
- Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A garbage can model of organizational choice. *Administrative Science Quarterly*, 17(1), 1–25.
- Cuban, L. (1988). *The managerial imperative and the practice of leadership in schools*. Albany: State University of New York Press.
- Cyert, R. M., & March, J. G. (1963). *A behavioral theory of the firm*. Englewood Cliffs, NJ: Prentice-Hall.
- Daft, R. L., & Weick, K. E. (1984). Toward a model of organizations as interpretation systems. *Academy of Management Review*, 9(2), 284–295.
- Datnow, A., Hubbard, L., & Mehan, H. (2002). *Extending educational reform: From one school to many*. London: Routledge/Falmer.
- Datnow, A., Park, V., & Wohlstetter, P. (2007). *Achieving with data: How high-performing school systems use data to improve instruction for elementary students*. Los Angeles: Center on Educational Governance.
- Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research*, 72(3), 433–479.
- Desimone, L., Porter, A. C., Birman, B. F., Garet, M. S., & Yoon, K. S. (2002). How do district management and implementation strategies relate to the quality of the professional development that districts provide to teachers? *Teacher College Record*, 104(7), 1265–1312.
- DuFour, R., & Eaker, R. (1998). *Professional learning communities at work: Best practices for enhancing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Earl, L., & Katz, S. (2002). Leading schools in a data-rich world. In K. Leithwood & P. Hallinger (Eds.), *Second international handbook of educational leadership and administration* (pp. 1003–1022). New York: Springer.

- Elmore, R. F., & Birney, D. (1997). *School variation and systemic instructional improvement in Community School District #2, New York City*. Unpublished manuscript. Retrieved February 14, 2008, from <http://www.lrdc.pitt.edu/hplc/Publications/School%20Variation.pdf>
- Floden, R. E., Goertz, M. E., & O'Day, J. (1995). Capacity building in systemic reform. *Phi Delta Kappan*, 77(1), 19–21.
- Firestone, W. A., & Hirsh, L. S. (2005). *A formative evaluation of the implementation of New Jersey's professional development requirements for teachers: Year five*. New Brunswick, NJ: Rutgers University, Center for Educational Policy Analysis.
- Fuhrman, S. H. (1993). The politics of coherence. In S. H. Fuhrman (Ed.), *Designing coherent education policy: Improving the system* (pp. 1–34). San Francisco: Jossey-Bass.
- Fuhrman, S. H., & Massell, D. (1992). *Issues and strategies in systemic reform*. Philadelphia: Consortium for Policy Research in Education.
- Fullan, M. (2005). *Leadership & sustainability. System thinkers in action*. Thousand Oaks, CA: Corwin.
- Gambone, M. A., Klem, A. M., Moore, W. P., & Summers, J. A. (2002). *First things first: Creating the conditions & capacity for community-wide reform in an urban school district*. New York: Manpower Demonstration Research Corporation.
- Glennan, T. K., Bodilly, S. J., Galegher, J. R., & Kerr, K. A. (2004). *Expanding the reach of education reforms*. Santa Monica, CA: RAND Corporation.
- Goodlad, J. (1999). Individuality, commonality, and curricular practice. In *Issues in curriculum: Selected chapters from NSSE yearbooks* (98th yearbook of the National Society for the Study of Education). (pp. 27–48), Chicago: University of Chicago Press.
- Gross, B., Kirst, M., Holland, D., & Luschei, T. (2005). Got you under my spell? How accountability policy is changing and not changing decision making in high schools. In B. Gross & M. Goertz (Eds.), *Holding high hopes: How high schools respond to state accountability policies* (CPRE Research Report Series RR-056, pp. 43–79). Philadelphia, PA: Consortium for Policy Research in Education.
- Herman, J. L., & Haertel, E. H. (2005). *Uses and misuses of data for educational accountability and improvement* (104th Yearbook of the National Society for the Study of Education). Malden, MA: Blackwell.
- Heshmati<sup>1</sup>, The United Nations University, Wider, A. (2003). Productivity growth, efficiency and outsourcing in manufacturing and service industries. *Journal of Economic Surveys*, 17(1), 79–112.
- Hightower, A. M., Knapp, M. S., Marsh, J. A., & McLaughlin, M. W. (2002). *School districts and instructional renewal*. New York: Teachers College Press.
- Honig, M. I. (2004). The New Middle Management: Intermediary Organizations in Education Policy Implementation. *Educational Evaluation and Policy Analysis*, 26(1), 65–87.
- Kahle, J. B. (1997). Systemic reform: Challenges and changes. *Science Educator*, 6(1), 1–6.
- King, J. A. (2002). Building the evaluation capacity of a school district. *New Direction for Evaluation*, 3, 63–80.
- Knapp, M. S., Zucker, A. A., Adelman, N. E., & St. John, M. (1991). *The Eisenhower Mathematics and Science Education Program: An enabling resource for reform* (Summary Rep.). Menlo Park, CA: SRI International.
- Kronley, R. A., & Handley, C. (2003). *Reforming relationships: School districts, external organizations, and systemic change* (Report prepared for School Communities that Work, A national task force on the future of urban districts). Providence, RI: Annenberg Institute for School Reform at Brown University.
- Leithwood, K. A. (1995). *Effective school district leadership: Transforming politics into education*. New York: SUNY Press.
- Levinthal, D., & March, J. G. (1981). A model of adaptive organizational search. *Journal of Economic Behavior & Organization*, 2(4), 307–333

- Lieberman, A., & Wood, D. R. (2003). *Inside the National Writing Project: Connecting network learning and classroom teaching*. New York: Teachers College Press.
- Mac Iver, M. A., & Farley, E. (2003). *Bringing the district back in: The role of the central office in improving instruction and student achievement*. Baltimore: Center for Research on the Education of Students Placed At Risk.
- Marsh, J. A., Kerr, K. A., Ikemoto, G. S., Darilek, H., Suttorp, M., Zimmer, R. W., et al. (2005). *The role of districts in fostering instructional improvement*. Santa Monica, CA: RAND Corporation.
- Massell, D. (2000). *The district role in building capacity* (CPRE Policy Brief). Philadelphia: Consortium for Policy Research in Education.
- Miles, K. H., Odden, A., Fermanich, M., & Archibald, S. (2005). *Inside the black box: School district spending on professional development in education*. Available from [http://www.wallacefoundation.org/WF/ELAN/TR/KnowledgeCategories/ImprovingConditions/UseOfFinancialResources/inside\\_the\\_black\\_box.htm](http://www.wallacefoundation.org/WF/ELAN/TR/KnowledgeCategories/ImprovingConditions/UseOfFinancialResources/inside_the_black_box.htm)
- Mosteller, F., & Boruch, R. (Eds.). (2002). *Evidence matters: Randomized trials in education research*. Washington, DC: Brookings Institution Press.
- Payzant, T. W., & Gardner, M. (1994). Changing roles and responsibilities in a restructuring school district. *NASSP Bulletin*, 78(560), 8–17.
- Petrides, L. A., & Guiney, S. Z. (2002). Knowledge management for school leaders: An ecological framework for thinking schools. *Teachers College Record*, 104(8), 1702–1717.
- Quint, J., Bloom, H. S., Black, A. R., Stephens, L. & Akey, T. M. (2005). *The challenge of scaling up educational reform: Findings and lessons from First Things First Final Report*. New York: Manpower Demonstration Research Corporation.
- Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (2003). *Evaluation: A systematic approach*. Thousand Oaks, CA: Sage.
- Schön, D. A. & McDonald, J. P. (1998). *Doing what you mean to do in school reform*. Providence, RI: Brown University.
- Senge, P. (1990). *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday.
- Simon, H. A. (1979). Rational decision making in business organizations. *The American Economic Review*, 69(4), 493–513.
- Smith, M. S., & O'Day, J. (1991). Systemic school reform. In S. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing* (pp. 233–267). New York: Falmer.
- Snipes, J., Doolittle, F., & Herlihy, C. (2002). *Foundations for success: Case studies of how urban school systems improve student achievement*. Washington, DC: Council of Great City Schools.
- Spillane, J. P. (1996). School districts matter: Local education authorities and state instructional policy. *Educational Policy*, 10(1), 63–87.
- Spillane, J. P., & Thompson, C. L. (1997). Reconstructing conceptions of local capacity. The local education agency's capacity for ambitious instructional reform. *Educational Evaluation and Policy Analysis*, 19(2), 185–203.
- Stuart, T., Hoang, H., & Hybels, R. (1999). Interorganizational Endorsements and the Performance of Entrepreneurial Ventures. *Administrative Science Quarterly*, 44, 315–349.
- Supovitz, J. A. (2006). *The case for district-based reform: Leading, building, and sustaining school improvement*. Cambridge, MA: Harvard Education Press.
- Supovitz, J. A. (2007). *Organizational learning and instructional improvement*. Paper presented at the annual conference of the American Educational Research Association, Chicago.
- Supovitz, J. A., & Klein, V. (2003). *Mapping a course for improved student learning: How innovative schools systematically use student performance data to guide improvement*. Philadelphia: Consortium for Policy Research in Education.
- Supovitz, J. A., & Weathers, J. (2004). *Dashboard lights: Monitoring implementation of district instructional reform strategies*. Philadelphia: Consortium for Policy Research in Education.

- Thomas, I. K. (2004). *Professional development for school leaders*. Washington, DC: American Association of Colleges for Teacher Education.
- Togneri, W., & Anderson, S. E. (2003). *Beyond islands of excellence: What districts can do to improve instruction and achievement in all schools*. Washington, DC: Association for Supervision and Curriculum Development.
- Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Cambridge, MA: Harvard University Press
- Tyler, R. W. (1988). Specific approaches to curriculum development. In J. R. Gress (Ed.), *Curriculum: An introduction to the field* (pp. 219–234). Berkeley, CA: McCutchan.
- Wayman, J. C., Stringfield, S., & Millard, M. O. (2004). *Software for disaggregating and reporting student data: Moving beyond "No Child Left Behind" to inform classroom practice*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.



# INSTRUCTIONS FOR AUTHORS

**Content:** *Peabody Journal of Education (PJE)* publishes quarterly symposia in the broad area of education, including but not limited to topics related to formal institutions serving students in early childhood, preschool, primary, elementary, intermediate, secondary, postsecondary, and tertiary education. The scope of the journal includes special kinds of educational institutions, such as those providing vocational training or the schooling for students with disabilities. PJE also welcomes immediate framework of institutions, and education matters that are important to nations outside the United States. Finally, it includes topics that are linked to the social and organizational context in which formal and informal education take place. The Editor cooperates with groups of scholars to present multifaceted, integrated expositions of important topics. A given issue of PJE may contain contributions from social scientists, historians, philosophers, attorneys, practitioners, and policymakers. Unsolicited proposals for special issues - including designation of participating scholars and an outline of articles - will be accepted for review. Additionally, the Editor cooperates with Editorial Board members to identify potential topics, guest editors, and contributors. PJE has the flexibility to consider publishing monographs or a series focused on particular lines of inquiry. In all cases, the Editor and the Editorial Board will ensure that each issue is carefully reviewed and its articles will comprise a high-quality contribution to understanding and practice.

**Manuscript Preparation:** To submit a prospectus for a special issue, note submission requirements presented here and contact the Coordinating Editor, Michele Thompson, for complete details (address below). Prepare manuscripts according to the *Publication Manual of the American Psychological Association* (5th ed., 2001; available from APA Order Department, P.O. Box 2710, Hyattsville, MD 20874), especially with regard to reference lists and text citations. Follow Guidelines to Reduce Bias in Language (*APA Manual*, pp. 61–87). Manuscripts should include the theme title and rationale for the theme. Using 8.5 × 11-in., nonsmear paper and 1-in. margins, type all components double spaced and in the following order: title page (p. 1), abstract (p. 2), text (including quotations), references, appendices, footnotes, tables, and figure captions. On page 1, type article title, author name(s) and affiliation(s), running head (abbreviated title, no more than 45 characters and spaces), author notes and acknowledgments, and name and address of person to whom requests for reprints should be addressed; on page 2, type an abstract of no more than 150 to 200 words. Indent all paragraphs. Use footnotes sparingly. Attach photocopies of all figures. Number all manuscript pages (including photocopies of figures).

**Cover Letter, Permissions, Credit Lines:** In a cover letter, include the contact author's complete mailing address, e-mail address, and telephone and fax numbers. State that the manuscript includes only original material that has not been published and that is not under review for publication elsewhere. Authors are responsible for all statements made in their work and for obtaining permission to reprint or adapt a copyrighted table or figure or to quote at length from a copyrighted work. Authors should write to original author(s) and original publisher to see if permission is required and to request nonexclusive world rights in all languages to use the material in the current article and in future editions. Include copies of all permissions and credit lines with the manuscript. (See p. 175 of the *APA Manual* for samples of credit lines.)

**Manuscript and Computer Disk Submissions:** Prepare manuscript on a word processor and submit three (3) high-quality printouts and a disk to the Coordinating Editor, Michele Thompson, *Peabody Journal of Education*, Department of Leadership, Policy, and Organizations, Peabody College of Vanderbilt University, Peabody #41, 230 Appleton Place, Nashville, TN 37203-5721. The disk should contain two files - word processor and ASCII versions of the manuscript - and file content must match the printed manuscript exactly, or there will be a delay in publication. Manuscripts and disks are not returned.

**Accepted Manuscripts:** Authors are strongly encouraged to submit manuscripts on disk. The disk should be prepared using MS Word or WordPerfect and should be clearly labeled with the authors' names, file name, and software program. A hardcopy printout that exactly matches the disk must be supplied. Each manuscript must be accompanied by a statement that it has not been published elsewhere and that it has not been submitted simultaneously for publication elsewhere. Authors are responsible for obtaining permission to reproduce copyrighted material from other sources and are required to sign an agreement for the transfer of copyright to the publisher. All accepted manuscripts, artwork, and photographs become the property of the publisher.

All parts of the manuscript should be typewritten, double-spaced, with margins of at least one inch on all sides. Number manuscript pages consecutively throughout the paper. Authors should also supply a shortened version of the title suitable for the running head, not exceeding 50 character spaces. Each article should be summarized in an abstract of not more than 100 words. Avoid abbreviations, diagrams, and reference to the text in the abstract.

**Illustrations:** Illustrations submitted (line drawings, halftones, photos, photomicrographs, etc.) should be clean originals or digital files. Digital files are recommended for highest quality reproduction and should follow these guidelines:

- 300 dpi or higher
- Sized to fit on journal page
- EPS, TIFF, or PSD format only
- Submitted as separate files, not embedded in text files

**Color Reproduction.** Color illustrations will be considered for publication; however, the author will be required to bear the full cost involved in color art reproduction\*. Color art can be purchased for online only conversion and reproduction or for print + online reproduction. Color reprints can only be ordered if print + online reproduction costs are paid. Rates for color art reproduction are:

**Online Only Reproduction:** \$225 for the first page of color; \$100 for each of the next three pages of color. A maximum charge of \$525 applies. **Print + Online Reproduction:** \$900 for the first page of color; \$450 per page for the next three pages of color. A custom quote will be provided for articles with more than 4 pages of color. \*Color art will be reproduced in color in the print and online publication at no additional costs to the author, if recommended by the Editor-in-Chief.

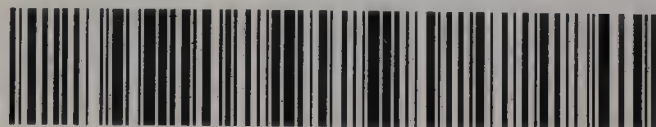
**Tables and Figures.** Tables and figures (illustrations) should not be embedded in the text, but should be included as separate sheets or files. A short descriptive title should appear above each table with a clear legend and any footnotes suitably identified below. All units must be included. Figures should be completely labeled, taking into account necessary size reduction. Captions should be typed, double-spaced, on a separate sheet. All original figures should be clearly marked in pencil on the reverse side with the number, author's name, and top edge indicated.

**Proofs and Reprints.** Page proofs are sent to the designated author using Taylor & Francis' Central Article Tracking System (CATS). They must be carefully checked and returned within 48 hours of receipt. Reprints of individual articles are available for order at the time authors review page proofs. A discount on reprints is available to authors who order before print publication.

# *Peabody Journal of Education*

*Volume 83, Number 3, 2008*

Introduction to the Special Issue on Scaling Up Teaching and Learning Improvement in Urban Districts: The Promises and Pitfalls of External Assistance Providers <i>Amanda Datnow and Meredith I. Honig</i>	323
Adaptive Assistance for Learning Improvement Efforts: The Case of the Institute for Learning <i>Meredith I. Honig and Gina S. Ikemoto</i>	328
Authority, Status, and the Dynamics of Insider–Outsider Partnerships at the District Level <i>Cynthia E. Coburn, Soung Bae, and Erica O. Turner</i>	364
Collaborative Assistance in a Highly Prescribed School Reform Model: The Case of Success for All <i>Vicki Park and Amanda Datnow</i>	400
Assistance and Accountability in Externally Managed Schools: The Case of Edison Schools, Inc. <i>Julie Marsh, Laura Hamilton and Brian Gill</i>	423
Melding Internal and External Support for School Improvement: How the District Role Changes When Working Closely With External Instructional Support Providers <i>Jonathan Supovitz</i>	459



0161956X(2008)83(3)

Volume 83 Issue 4 ISSN 0161-956X 2008

# *Peabody Journal of Education*

**CURRENT YR/VOL**

**Marygrove College Library  
8425 West McNichols Road  
Detroit, MI 48221**

*Published in cooperation with the Peabody College of Vanderbilt University*

**ROUTLEDGE**



**Routledge**  
Taylor & Francis Group



# PEABODY JOURNAL OF EDUCATION

Camilla P. Benbow, Dean, Peabody College  
James W. Guthrie, Editor  
Michele Thompson, Coordinating Editor  
Deborah M. Enright, Associate Editor

## EDITORIAL ADVISORY BOARD

Darlyne Bailey, *University of Minnesota*  
R. Dale Ballou, *Peabody College of Vanderbilt University*  
Sir Clive Booth, *Oxford Brookes University, England*  
Mary M. Brabeck, *New York University*  
Mark Bray, *The University of Hong Kong, PRC*  
Louis A. Castenell, Jr., *The University of Georgia*  
David Y. Chen, *Chongqing Technology and Business University (CTBU)*  
Mike Dishman, *McLocklin, Murphy & Dishman, L.L.P.*  
Lora Cohen-Vogel, *Florida State University*  
Joseph P. Du Cette, *Temple University*  
Victor Firsov, *Russian Academy of Teachers Retraining, Russia*  
Juan Froemel, *LLECE/UNESCO, Chile*  
Louis M. Gomez, *Northwestern University*  
Gerardo M. Gonzalez, *Indiana University*

David W. Gordon, *Elk Grove Unified School District, California*  
Eric A. Hanushek, *Hoover Institution, Stanford University*  
Stephen P. Heyneman, *Peabody College of Vanderbilt University*  
Phillip W. Jones, *The University of Sydney, Australia*  
Kazuo Kuroda, *Waseda University, Japan*  
James H. Lytle, *Trenton Public Schools, New Jersey*  
Michael K. McLendon, *Peabody College of Vanderbilt University*  
Ann L. Marcus, *New York University*  
Bill McKinney, *Region IV Education Service Center, Texas*  
Francois Orivel, *University of Bourgogne, France*  
Penelope L. Peterson, *Northwestern University*  
Michael J. Podgursky, *University of Missouri-Columbia*  
Robert Slavin, *Johns Hopkins University*  
Jandhylala B. G. Tilak, *National Institute of Education Planning and Administration, India*

---

**Abstracted and/or Indexed in:** PsycINFO/Psychological Abstracts; Linguistics and Language Behavior Abstracts; Contents Pages in Education; Sociological Abstracts; Education Index; Education Abstracts; Social Services Abstracts; EBSCOhost Products; Cabell's Directories; Family Index Database.

*Peabody Journal of Education* (ISSN: 0161-956X) is published quarterly in February, May, August, and November for a total of 4 issues per year by Taylor and Francis Group, LLC, 325 Chestnut Street, Suite 800, Philadelphia, PA 19106.

**US Postmaster:** Please send address changes to Peabody Journal of Education, Taylor & Francis Group, LLC, 325 Chestnut Street, Suite 800, Philadelphia, PA 19106.

### Annual Subscription, Volume 83, 2008

Print ISSN - 0161-956X, Online ISSN - 1532-7930

Institutional subscribers: US \$462, UK £277, €370

Personal subscribers: US \$58, UK £35, €46

Institutional and individual subscriptions include access to the online version of the journal. Institutional subscriptions include access for any number of concurrent users across a local area network. Individual subscriptions are single username/password only.

**Production and Advertising Office:** 325 Chestnut Street, Suite 800, Philadelphia, PA 19106. Tel: 215-625-8900, Fax: 215-625-8563. Production Editor: Michelle McCarthy

### Subscription offices

**USA/North America:** Taylor and Francis Group, LLC, 325 Chestnut Street, Suite 800, Philadelphia, PA 19106. Tel: 215-625-8900, Fax: 215-625-2940.

**UK/Europe:** Taylor and Francis Customer Services, Sheepen Place, Colchester, Essex CO3 3LP, UK. Tel: +44 (0)-20-7017-5544, Fax: +44 (0)-20-7017-5198.

For a complete guide to Taylor and Francis Group, LLC's journal and book publishing programs, visit our website: [www.taylorandfrancis.com](http://www.taylorandfrancis.com)

**Copyright** © 2008 Taylor and Francis Group, LLC. All rights reserved. No part of this publication may be reproduced, stored, transmitted, or disseminated in any form or by any means without prior written permission from Taylor and Francis Group, LLC. Taylor and Francis Group, LLC grants authorization for individuals to photocopy copyright material for private research use. The publisher assumes no responsibility for any statements of fact or opinion expressed in the published papers. The appearance of advertising in this journal does not constitute an endorsement or approval by the publisher, the editor, or the editorial board of the quality or value of the product advertised or of the claims made for it by its manufacturer.

**Permissions.** For further information, please visit: <http://www.tandf.co.uk/journals/permissions.asp>



# Introduction to the MSP-PE Special Issue on Math and Science Partnership Program: A First Comprehensive Evaluation

Kenneth K. Wong  
*Brown University*

Robert K. Yin  
*COSMOS Corporation*

Patricia S. Moyer-Packenham  
*Utah State University*

Jennifer Scherer  
*COSMOS Corporation*

The United States faces significant challenges in the fields of science, technology, engineering, and mathematics (often collectively referred to as STEM). Numerous reports from governmental, scientific, and civic communities have raised concerns over the quality of STEM education at all levels of the educational system, the shortage in the STEM labor force, and the decreasing competitiveness of student performance in STEM fields at the international level.

---

The coeditors of this special issue acknowledge the support provided by Darci Terrell of COSMOS Corporation in the preparation of this journal. The authors thank the National Science Foundation staff, members of the evaluation's advisory board, and a number of other experts for their comments on earlier drafts of the studies. The authors also gratefully acknowledge the work of the Math and Science Partnership Program participants. Any opinions, findings, conclusions, and recommendations expressed in this special issue are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Correspondence should be sent to Kenneth K. Wong, Brown University, Education Department, Box 1938, Providence, RI 02912. E-mail: [Kenneth.Wong@brown.edu](mailto:Kenneth.Wong@brown.edu)

One indicator of the challenges lies in international comparisons of student performance in math and science. The 2003 Trends in International Mathematics & Science Study, conducted by the National Center for Education Statistics of the U.S. Department of Education, ranked the United States sixth in fourth grade and ninth in eighth grade among industrialized nations in student performance in science (International Association for the Evaluation of Educational Achievement, 2003; Martin, Mullis, Gonzalez, & Chrostowski, 2004). Furthermore, according to the 2003 Programme for International Student Assessment, an initiative of the Organisation for Economic Cooperation and Development (OECD) which assesses 15-year-olds' problem-solving performances on various subjects, the United States scored below the average performance for the OECD countries (National Academy of Sciences, 2007).

In light of these mixed performance records, the U.S. Congress has authorized several initiatives. Among the major strategies to address these concerns in STEM fields is the Math and Science Partnership (MSP) Program, a major national initiative funded by the National Science Foundation (NSF). As the NSF's (2001) original solicitation in 2002 stated, the MSP Program "seeks to improve student outcomes in high-quality mathematics and science for all students, at all pre-K-12 levels." At the same time, the program promotes research and development in STEM. Toward these multiple objectives, the program requires one or more Institutions of Higher Education (IHEs) to partner with K-12 public school districts to improve STEM activities. Since 2002, the MSP Program has awarded four cohorts of MSP grantees. The first three cohorts totaled 48 MSP partnerships and 29 related awards in 2002, 2003, and 2004. The work by the 48 MSPs is the subject of the studies highlighted in this issue.

Given the prominence of the MSP Program, the NSF has commissioned a multidisciplinary team of researchers from COSMOS Corporation, Brown University, and George Mason University to conduct a multiyear external evaluation. The collection of studies in this special issue represents a coordinated and initial effort to evaluate the design and implementation, as well as some of the effects of the MSP Program. Taken as a whole, the research team maintains a comprehensive pool of disciplinary knowledge, including mathematics, chemistry, biology, physics, engineering, education, economics, political science, statistics, and policy and program evaluation. Team members engage in a number of substudies that adopt different research designs and methods that range from econometric, psychometric, to qualitative and documentary analyses.

Led by COSMOS Corporation, the research team recognizes several design realities. The MSP Program consists of a set of separately funded projects. Each project was independently reviewed and approved as part of NSF's rigorous peer review process. In this regard, the program attracted applicants who were likely to be experienced in organizing STEM activities that connect IHEs and school districts. In operational terms, the program is defined by its awardees and the

specific context within which each project is situated. Although some MSPs invest in enhancing the quality of STEM activities at the university level, others focus on in-service activities on a particular STEM subject in a specific grade span in a cluster of public schools. The MSP Program therefore cannot be considered a homogenous effort that might, for instance, follow any singular research design, such as a randomized control experiment. Indeed, the MSP projects, themselves, employ an array of evaluative and research methods to study their varied strategies.

To study such a complex program that maintains multiple sites, institutions, foci, and relationships, the evaluation team has adopted a comprehensive evaluation agenda that spans K-20. In his overview of the evaluation effort, Robert K. Yin highlights that the challenge of the program evaluation is nevertheless to consider the MSP Program as a whole and not to assess any of the awards individually. His study traces the rationale behind a multi-institutional framework that covers a series of pathways in the K-20 span of mathematics and science education. For example, high school graduates may proceed to undergraduate and graduate careers, including the teaching profession that instructs the next generation on STEM fields. This systems approach calls for a series of substudies that collectively address the multifaceted interorganizational and intraorganizational relationships in the MSP Program. The early substudies are then reported in the ensuing studies in this journal.

Three studies examine the challenge of teacher quality and supply in math and science. In "A Review of the Literature on Mathematics and Science Teacher Quality," Johnna Bolyard and Patricia S. Moyer-Packenham synthesize approximately 150 studies on teacher quality and student outcomes in mathematics and science. At the secondary level, the authors found a generally positive relationship between teacher subject matter knowledge and pedagogical training and student achievement. However, at the elementary level, the relationship seems to be inconclusive. This may be because of the observation that "elementary teachers are usually generalist and their credentials reflect this status." These findings are likely to have broad implications on teacher training.

Using econometric methods, John Tyler and Svetla Vitanova examine the relationship between the MSP Program and the supply of certified teachers in mathematics. In recent years, numerous studies have identified the shortage of certified math teachers as an important factor in the lack of academic progress in mathematics. In "Does MSP Participation Increase the Supply of Math Teachers? Developing and Testing an Analytic Model," the authors propose a set of analytic parameters in estimating the extent to which the MSP Program may address this challenge of math teacher shortage. At issue is whether the MSPs can increase teacher supply given existing constraints, including districts' use of uncertified teachers, lack of flexibility in using differential salaries to attract teachers in math and science, and the value on salaries that potential teachers (or college graduates)

place on compensation in the labor market. Using Texas's three MSPs for illustrative purposes, Tyler and Vitanova argue for the reasonableness of their developed model in estimating the MSP Program effect on teacher supply.

Patricia S. Moyer-Packenham, Johnna Bolyard, Anastasia Kitsantas, and Hana Oh examine the ways in which grantees in the MSP Program document teacher quality in math and science fields. The research team analyzed 123 annual and evaluation reports, in addition to awardees' Web sites, publications, and presentations. Based on an extensive documentary analysis of 48 MSP-funded projects, the research team found that the awardees have relied on externally designed surveys and observations to define teacher quality and characteristics, including teacher beliefs and subject knowledge. The awardees' focus on these kinds of teacher characteristics did not come as a surprise, as they are connected to student achievement. Although awardees' documents show their understanding on the complexity of teaching, locally designed instruments often lack psychometric information.

Closely connected to teacher supply and quality is the delivery of curriculum, an issue addressed in "Mathematics Curriculum Systems: Models for Analysis of Curricular Innovation and Development." In this study, Margret A. Hjalmarson applies three models to analyze and categorize curriculum systems in the MSP Program sites. The three analytical perspectives are not meant to be mutually exclusive but instead provide different lenses on the curriculum foci. First, the content-based model directs our attention to the mathematics a student should know. It enables us to investigate how students engage in learning and how teachers address standards-based objectives. Second, the pedagogically based approach illuminates the instructional methods used to engage the students. Particularly relevant are teachers' belief systems, mathematical knowledge, skills development, and interpretative practices. Third, the learner-centered perspective pays particular attention to learner-related goals and the ways teachers provide support for accomplishing these goals. This perspective enables us to consider the learning gaps among student subgroups.

To be sure, curricular and other activities in the MSP sites are situated in the broader context of partnerships between IHEs and school districts. In "A Review of Instruments to Evaluate Partnerships in Math and Science Education," Jennifer Scherer argues the importance of conducting self-evaluation as part of the ongoing effort to improve the work of partnerships. The author conducts a careful synthesis of the literature on self-evaluation and the evaluation instruments across various fields in human, social, and education services. This comprehensive review shows that there are a number of useful assessment instruments that measure the context, structure, capacity, and the intergovernmental and intraorganizational conditions of partnerships. The article observes the utility in making use of different aspects of these existing instruments to address the needs of the MSPs. In other words, there are many existing tools available for self-assessment purposes.



Two studies address the issue of student achievement from different analytical perspectives. In “Initial Trends in MSP-Related Changes in Student Achievement with MIS Data,” Dimitar Dimitrov uses a within-group design and examines the relationship between the degree of MSP Program participation and student achievement. The annual survey of K-12 districts in the MSP Program for the first three program years provided the data for school and teacher participation as well as the school identification for gathering student achievement data. During the first three program years, the MSP Program’s participating schools show overall improvement in math and science proficiency. In examining teacher participation in MSP activities, Dimitrov observed a positive relationship between schools’ targeted teacher participation in MSP-related activities and student proficiency in math and science at the elementary and high school levels. No observable relationship is found for middle schools. Because this article uses a within-group design, it does not include a control group for the analysis. The latter is the focus of the next study.

Kenneth Wong and Ted Socha employ a comparative approach on student achievement. Their pilot study proposes a set of analytical steps for comparing schools that participate in the MSP Program and their nonparticipating peers in the same state. The study focuses on a sample of participating schools in one MSP in one state as identified by the annual survey of the K-12 districts in the MSP Program. The nonparticipating schools were systematically matched with the program’s participating schools on eight demographic variables to form a comparison group. Student performance data come from publicly accessible school-level data that the research team retrieved from the state’s department of education Web site for 2002–03 through 2004–05, as well as data available from the National Center for Education Statistics’ Common Core of Data. This article offers detailed documentation on how to operationalize two matching methods for comparative purposes. The article concludes that carefully executed matching methods are promising for large-scale comparative analysis on the effects of the MSP Program across different states.

Finally, Robert K. Yin, Daryl Chubin, and Edward Hackett investigate the complex issue of innovative activities in the broader context of the MSP Program as an education research and development (R&D) effort. In “Discovering ‘What’s Innovative’: The Challenge of Evaluating Education R&D Efforts,” the research team argues that the MSP Program can be assessed by contributions made to new ideas and practices in education. Because all R&D activities can be described in terms of one or more of four processes—namely, uncovering, inventing, explaining, and substantiating—the evaluation team can focus on monitoring innovative outcomes by examining evidence about the four processes in the MSP Program.

The studies in this special issue are based on analyses completed in 2006. The evaluation has continued to evolve with additional analyses and new findings for future publications. This evaluation is supported by the NSF’s MSP Program

through contract no. EHR-0456995: “Math and Science Partnership Program Evaluation.” Since 2007, Bernice Anderson, Ed.D., Senior Advisor for Evaluation, Directorate for Education and Human Resources, has served as the NSF Program Officer. The MSP-PE is led by COSMOS Corporation in current partnership with George Mason University and Brown University. Robert K. Yin (COSMOS) serves as Principal Investigator and Jennifer Scherer (COSMOS) serves as one of three Co-Principal Investigators. Additional Co-Principal Investigators and their collaborating institutions (including discipline departments and math centers) are Patricia Moyer-Packenham (Utah State University) and Kenneth Wong (Brown). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors do not necessarily reflect the views of the NSF.

REFERENCES

International Association for the Evaluation of Educational Achievement. (2003). *Trends in International Mathematics and Science Study (TIMSS)*. Amsterdam: Author.

Martin, M. O., Mullis, V. S., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 International Science Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

National Academy of Sciences. (2007). *Rising above the gathering storm*. Washington, DC: The National Academies Press.

National Science Foundation. (2001). *Math and Science Partnership (MSP), Program Solicitation*. Arlington, VA: Author.

APPENDIX

Glossary of Terms and Abbreviations for the Math and Science Partnership Program Evaluation (MSP-PE)

<b>AP Program</b>	Advanced Placement Program
<b>ED</b>	U.S. Department of Education
<b>ED-MSP</b>	Mathematics and Science Partnerships program administered by the U.S. Department of Education; a counterpart to NSF’s MSP Program
<b>IHE</b>	Institution of higher education
<b>LEA and SEA</b>	Local education agency and state education agency
<b>MAT</b>	Master of Arts in Teaching
<b>M/S or M&amp;S</b>	Math and science
<b>MSP Program or NSF-MSP</b>	NSF’s Math and Science Partnership Program.

<b>MSP-MIS</b>	Math and Science Partnership (Program's) Management Information System, to obtain annual data from each MSP-funded project
<b>MSPnet</b>	MSPnet (the Math Science Partnership Network) provides the MSP program with a web-based, interactive electronic community ( <a href="http://www.mspnet.org">www.mspnet.org</a> )
<b>MSP-PE</b>	Math and Science Partnership Program Evaluation
<b>MSPs or MSP awardees</b>	Math and Science Partnership awardees funded by the National Science Foundation under the MSP Program
<b>NAS</b>	National Academy of Sciences
<b>NCLB</b>	The No Child Left Behind Act signed into law in January 2002
<b>NSB</b>	National Science Board
<b>NSF</b>	National Science Foundation
<b>OMB clearance</b>	Office of Management and Budget, an agency of the executive branch of the federal government; OMB clearance is required to collect data from 10 or more individuals using a standardized data collection instrument
<b>PD</b>	Professional development
<b>PIs or co-PIs</b>	Principal investigators or co-principal investigators
<b>Pre-K-12</b>	Encompasses pre-Kindergarten, Kindergarten, and grades 1-12
<b>R&amp;D</b>	Research and development
<b>RETA</b>	Research, Evaluation, and Technical Assistance
<b>STEM (education)</b>	Science, technology, engineering, and mathematics (education)
<b>TA</b>	Technical Assistance
<b>TIMSS</b>	Trends in International Mathematics and Science Study

# The Math and Science Partnership Program Evaluation: Overview of the First Two Years

Robert K. Yin  
*COSMOS Corporation*

This study describes the Math and Science Partnership Program Evaluation (MSP-PE) during the project's first two years and provides the evaluation framework being used to assess the National Science Foundation's MSP Program. The study conveys the MSP-PE's ongoing design and implementation. To show how they reflect the nature of the MSP Program, the study addresses the following questions: (a) What are the MSP Program's main themes? (b) What kinds of activities have the program's awardees been putting into place? (c) What are the awardees doing to assess K-12 student achievement outcomes? and given the preceding conditions, (d) What is the framework and design for the MSP-PE? The study shows how the framework and the emerging evaluation derive from the program's main themes and its early activities, also giving readers a glimpse of the program's activities. The study traces the rationale behind a multi-institutional framework that covers a series of pathways in the K-20 span of mathematics and science education. This systems approach calls for a series of substudies that collectively address the multifaceted interorganizational and intraorganizational relationships in the MSP Program. The evaluation's framework provides a unifying scope for the series of substudies—all of which have been undertaken as part of the MSP-PE. Some of MSP-PE's early substudies are contained in this special issue.

## THE ROLE OF K-12 MATHEMATICS AND SCIENCE EDUCATION IN STRENGTHENING SCIENCE, MATHEMATICS, AND ENGINEERING

This study presents the evaluation framework being used to assess a major national initiative, the National Science Foundation's (NSF's) Math and Science



Partnership (MSP) Program. The study shows how the framework and the emerging evaluation derive from the program's main themes and its early activities, also giving readers a glimpse of those activities.

## THE MSP PROGRAM IN THE CONTEXT OF NATIONAL ATTENTION DEVOTED TO K-12 MATHEMATICS AND SCIENCE EDUCATION

Stepping back for a moment, the MSP Program has been taking place as part of a continuing focus on the importance of "K-12"<sup>1</sup> mathematics and science education in this country. For NSF, the program is integral to its broader mission, helping the United States to maintain a position of eminence at the global frontier of "fundamental and transformative scientific research" and to sustain a "world class science and engineering (S&E) workforce"<sup>2</sup>—while also fostering the scientific literacy of all citizens (National Science Board [NSB], 2005, p. 3). The S&E workforce includes not only practicing scientists and engineers but also teachers (and in particular K-12 teachers) of mathematics and science. To serve current and future generations, the successful workforce must draw from students who have gained a strong mathematics and science education. The critical nature of the K-12 system arises from its positioning at the beginning of such education.

Reflecting concern over the needs of the S&E workforce, officials at the U.S. congressional audit agency (the U.S. General Accountability Office [GAO]) testified about S&E shortfalls before a congressional committee in May 2006 (GAO, 2006). For the S&E workforce in general, the officials noted that: Over a ten-year period, employment in science, technology, engineering, and mathematics (STEM) fields had risen more than the number of students graduating with STEM degrees (see Table 1). For K-12 mathematics and science teachers, the agency produced additional data regarding the importance of such teachers, as well as the rate of course completion in high school, as two critical influences that "affected students' success in and decisions about pursuing STEM fields" (GAO, 2006, p. 8). Among its top priorities, the agency recommended an increased university presence in pre-K-12 STEM education and a renewed commitment to graduate education (p. 9).

---

<sup>1</sup>The MSP Program is concerned with all primary and secondary school grades, pre-Kindergarten through Grade 12. For convenience's sake, "K-12" is used throughout to reference this range of grades rather than the more cumbersome "pre-K-12."

<sup>2</sup>These were the first two strategic priorities, respectively, in the NSB's vision statement for NSF. The third strategic priority was to build the nation's basic research capacity by making critical investments in infrastructure, including advanced instrumentation, facilities, cyberinfrastructure, and cutting-edge experimental capabilities (NSB, 2005).

TABLE 1  
Graduates and Employment In Science, Technology, Engineering, and Mathematics  
Graduates With STEM Degrees

1994-1995		2003-2004		Percent Change from 1994-1995 to 2003-2004
Number	As Percent of All Degrees Awarded	Number	As Percent of All Degrees Awarded	
519,000	32	578,000	27	No. of degrees awarded +11.4 Percentage of all degrees awarded -16.0
Employment Sector				Percent Change in Employment, 1994-2003
STEM Fields				+23.0
Non-STEM Fields				+17.0

Source: U.S. General Accountability Office. “Science, Technology, Engineering, and Mathematics Trends and the Role of Federal programs,” May 3, 2006.

Likewise, an expert group convened by the National Academy of Sciences (NAS), empanelled as the “Committee on Prospering in the Global Economy of the 21st Century,” also directed its attention to K-12 mathematics and science education. Among all of the initiatives the group considered when it concluded its work in the fall of 2005, the group recommended, first and foremost, that federal policy needed to increase America’s S&E talent pool by “vastly improving K-12 science and mathematics education” (NAS Committee on Science, Engineering, and Public Policy, 2007, p. 5). To implement this policy initiative, the panel’s first recommended action was to “annually recruit 10,000 science and mathematics teachers by awarding four-year scholarships and thereby educating 10 million minds” (NAS Committee on Science, Engineering, and Public Policy, 2007, p. 5).

The GAO and NAS examples, as well as others,<sup>3</sup> contribute to a highly visible backdrop within which the MSP Program has been operating.<sup>4</sup> The program

<sup>3</sup>For instance, strengthening the focus and funding to improve K-12 mathematics education in particular also appears as a priority in the legislation related to the American Competitiveness Act (e.g., Domestic Policy Council, 2006; Sroufe, 2006).

<sup>4</sup>The priority given to the MSP was sufficiently high that the NSF-MSP has a counterpart initiative—the “Mathematics and Science Partnerships”—at the U.S. Department of Education (ED-MSP). A U.S. House Committee report described the complementarity of the two initiatives as follows: Whereas NSF’s program is to fund “innovative programs to develop and establish new models of education reform, thereby remedying the lack of knowledge about math and science research,” ED’s program

TABLE 2  
Coverage of MSP Awards by MSP-PE (Reported in the MSP-MIS)

Type of MSP	Cohort				Total
	I (2002)	II (2003)	III (2004)	IV (2006)	
Comprehensive partnerships	6	6	0	0	12
Targeted partnerships	16	7	6	0	29
Institute partnerships	0	0	7	NA	7
RETA	6	11	5	NA	22
Total	28	24	18	NA	70

*Note.* Source: Author, and updates from NSF-MSP Program staff. The data do not include all awards made by the MSP Program. For instance, some awards were “design” awards only, and others ended early by mutual agreement. MSP = Math and Science Partnership; MSP-PE = Math and Science Partnership Program Evaluation; MIS = Management Information System; NA = not available; RETA = Research, Evaluation, and Technical Assistance.

consists of a portfolio of separately awarded (extramural) projects, most being supported for five years. To date, the program has made four rounds of awards since it first started in 2002 (see Table 2).

### The Evaluation of the MSP Program

The Math and Science Partnership Program Evaluation (MSP-PE) is the national, multisite evaluation of the MSP Program. The MSP-PE started in 2004, and its first year was devoted to design and planning. In addition, because of the necessary clearances, the evaluation’s own original fieldwork only started in the latter half of 2006.

The purpose of the study presented here is to convey the MSP-PE’s ongoing design and implementation. To show how they reflect the nature of the MSP Program, the study addresses four questions:

1. What are the MSP Program’s main themes?
2. What kinds of activities have the program’s awardees been putting into place?
3. What are the awardees doing to assess K-12 student achievement outcomes? and given the preceding conditions
4. What is the framework and design for the MSP-PE?

is aimed at “broadly implementing and disseminating new teaching materials, curricula, and training programs” (U.S. House of Representatives, 2003, p. 4). The ED-MSP is being separately evaluated, falling outside of the purview of the MSP-PE and hence of this study.

### *What Are the MSP Program's Main Themes?*

Several "root" documents define the MSP Program's main themes. The documents include the original authorizing language from Congress (P.L. 107-368) and five subsequent proposal solicitations issued by the program (listed in the references and also cited next). The language in these solicitations provides (a) background to the MSP Program, (b) statements about its mission, and (c) descriptions of relevant activities that might be supported. Collectively, the information from all these documents depicts the MSP Program in a multifaceted manner.

*Student outcomes.* First, the initial MSP solicitation starts by saying that the MSP Program "seeks to improve student outcomes in high-quality mathematics and science by all students, at all pre-K-12 levels" (NSF-02-061; NSF, 2001). This theme is echoed throughout all of the subsequent solicitations, which also identify "student achievement" as the outcome of greatest interest among the student outcomes. However, although the MSP Program focuses on student outcomes, this objective takes place within the broader context of "improving elementary and secondary mathematics and science *education* [emphasis added]" (P.L. 107-368). Improvements in education can result from related initiatives, such as increasing the quality, quantity, and diversity of K-12 teachers (NSF-02-061)—a workforce goal that suits well NSF's broader agency mission previously discussed—independent of any demonstration of student achievement outcomes.

*Partnerships between institutions of higher education (IHEs) and K-12 districts.* Second, the MSP Program fosters interorganizational partnerships, requiring one (or more) IHEs to be teamed with one (or more) K-12 school district(s). Together, these two institutions cover the entirety of the K-20 span of science and mathematics education. Without such partnerships, the pathway for entering the S&E workforce can be inefficient if not disjointed, even if students successfully negotiate their way through either system alone. Similarly, without such partnerships, consistency between what teachers need to know for K-12 classrooms, and what aspiring teachers are themselves taught in their undergraduate and graduate IHE courses, may assume an undesirable, more serendipitous nature.

This core, IHE-district(s) partnership also calls into play the "substantial engagement" of IHE discipline faculty in an MSP project. The engagement of such faculty "is considered one of the attributes that distinguishes the MSP program from other programs seeking to improve K-12 student outcomes in mathematics and science" (NSF-06-539; NSF, 2006). The resulting partnerships may cover IHE programs that produce candidates for K-12 teaching in mathematics and science; improve K-12 curricula, instruction, and assessment systems; develop technology to support instruction; or perform related functions. In this sense,



MSP builds on the Nation's dedication to improve mathematics and science education through support of partnerships that unite the efforts of local school districts with faculties of colleges and universities—especially disciplinary faculties in mathematics, science, and engineering—and with other stakeholders. (NSF-03-541; NSF, 2003b)

The other stakeholders can include family, community, and public and private organizations (including science centers and other “informal” science institutions as well as businesses and industry) who are part of the broader education community in most locales. From the perspective of evaluating the MSP Program, the breadth of the institutional span leads to the need for a multi-institutional, K-20 framework, and not just a single-institution, K-12 framework.

*Multiple Permissible Activities.* Third, within the K-20 span, the MSP Program may support many different kinds of activities. The authorizing language (P.L. 107-368) itself lists 13 possible activities, also giving NSF the discretion to support “any other activities” that will accomplish the goals of the program. Furthermore, the listed activities, such as “recruiting and preparing students for mathematics and science careers” (P.L. 107-368), are broad enough to themselves consist of multiple initiatives.

In this sense, the MSP Program does not impose a uniform set of activities on its awardees. Such flexibility befits the diversity of local education conditions within which the MSP's awardees are to operate, and the MSP Program's portfolio not surprisingly consists of a heterogeneous group of awarded projects. At the same time, the program brings the diverse array of activities under a program rubric that highlights *five key features* that awardees are to emulate: (a) being partnership driven; (b) striving for teacher quality, quantity, and diversity; (c) emphasizing challenging courses and curricula for students; (d) pursuing an evidence-based design and outcomes; and (e) seeking institutional change and sustainability (NSF-03-605; NSF, 2003a).

*A research and development (R&D) effort.* Fourth, the MSP Program positions itself as “a major research and development effort” (NSF-03-605). As such, the program occupies a dual niche. It is concerned with having an impact on large numbers of K-12 students (signalled by the repeated encouragement in the award solicitations to form partnerships or consortia that can serve 10,000 or more K-12 students—while noting that the size of the awards would be proportional to the number of students likely to be served—NSF-02-061 [NSF, 2001] and NSF-02-190 [NSF, 2002]). However, the program also is concerned with discovering new (and improved) ways of providing K-12 education (see Yin, Hackett, & Chubin, 2008/this issue).

The R&D theme was not explicitly present in the earliest proposal solicitations. Nevertheless, the MSP Program needs to be construed (and evaluated) both as an R&D program and as a program concerned with implementing change in existing K-12 systems.

### *What Kinds of Activities Have the MSP Program's Awardees Been Putting Into Place?*

The MSP Program began making awards before the start of the MSP-PE evaluation. This enabled the MSP-PE team to get an early glimpse of the activities actually being implemented by the awardees, with the information mainly coming from the awardees' annual reports to NSF.

*Describing MSP awards and their component "activities".* The initial description reflects the work of the 35 "comprehensive" and "targeted" MSP awards<sup>5</sup> (MSPs) funded by the MSP Program in its first two cohorts, starting in 2002 and 2003. Importantly, the compilation of activities recognized that every MSP has not necessarily been limited to a singular activity. The breadth of the MSP Program's mission, as well as the size of the awards (in some cases covering millions of dollars per year for each of five years), has led to a common situation whereby a single MSP may be undertaking two or more different activities.

For instance, one part of an MSP might be devoted to providing inservice training, or professional development, to existing K-12 teachers. Another part of the same MSP might simultaneously be strengthening a preservice program to encourage more undergraduate candidates to consider K-12 teaching careers in mathematics and science. The two activities would be independent to the extent that each exhibited

- Separate goals or objectives (even if related to those of other activities).

---

<sup>5</sup>The MSP Program distinguishes among four types of awards. Of the 53 awards still active from the first two cohorts, 35 were either "comprehensive" (an awardee covers the entire K-12 grade span) or "targeted" (an awardee covers a selected number of grades). Only 1 award fell into the third category of teacher institutes (an awardee focuses on teacher training institutes). The remaining 17 awards fell into the fourth category of "research, evaluation, or technical assistance" awards (an awardee chooses to study, support, provide tools for, or otherwise collaborate with one or more of the "comprehensive," "targeted," or "institute" awardees, and the collaboration also may include similar projects not funded directly by the NSF-MSP). Although the MSP-PE evaluation framework and design embrace all four types of awards, the description of the MSP activities in this study is limited to the first two types of awards only. The evaluation initially intends to focus separately on the other two types of awards, later bringing and synthesizing the lessons learned from all four types into a fuller comprehensive assessment of the entire MSP Program.

- Self-contained sets of participants, procedures, materials, and actions.
- A distinctive name, label, or other means of identification.
- Separately tracked funds and resources (even if part of a larger pool).
- A self-contained micro-organization consisting of staff, schedules, and other logistical details.

Conversely, two different activities might appear on the surface to be independent but in fact be part of the same, close-knit initiative. For instance, one part of an MSP might be devoted to strengthening the K-12 curriculum in mathematics and science. Another part might be providing professional development, but the professional development is limited to those K-12 teachers who are to implement the strengthened curriculum. Under this circumstance, what might at first have appeared to be two different activities would be considered two parts of a single activity.

Both kinds of situations exist among the MSPs. Yet the annual reports do not provide sufficient information to distinguish authoritatively between them. As a result, for the purpose of characterizing the 35 MSPs in the first two cohorts of the MSP portfolio, no assumption has been made about the relationship between or among activities within the same MSP. For the time being, where an MSP reports different activities, these have been treated as separate. (Ultimately, the evaluation team intends to use its own fieldwork to support any final clarification.)

*The distribution of activities, for 35 MSPs.* As expected, the review revealed a wide array of activities. Nevertheless, they could be clustered according to whether they tended, first, to be taking place (I) within a K-12 system; (II) within an IHE; or (III) with families, community, and public and private organizations. Within the K-12 and IHE systems, further “sub”clusters then identified whether the activity tended to emphasize: (A) working with students and classrooms or courses directly, (B) working with faculty and staff, or (C) working with institutional policies and structure.

Table 3 summarizes the distribution of the activities from the 35 MSPs according to these clusters and subclusters. The table shows 102 activities having been identified among the 35 MSPs. If each activity were truly separate, on average every MSP would consist of about 3 activities. However, many of these activities are likely to be found later to be parts of the same larger activity, so the actual number of activities lies somewhere between 35 and 102. Overall, and as might be expected from the thrust of the MSP Program, Table 3 shows that, among the three *clusters*, more of the activities were in the K-12 system, fewer were in the IHE system, and only a small portion dealt directly with families, community, and public and private organizations.

Table 3 also shows that the various MSP activities did not fall evenly among the *subclusters*. Much (36%) of the work of these 35 MSPs appears to involve the subcluster “working with K-12 teachers and staff” (Item IB in Table 3)—usually in some sort of inservice or professional development activity.

Conversely, although the awardees reported many activities in the IHE system, none of the 102 activities fell within the corresponding subcluster—“working with IHE faculty, administrators, or staff” (Item IIB in Table 3). Such an absence was not surprising, given the absence of such emphasis more generally in the IHE system (and hardly specific to the MSP Program). For example, any inservice or professional development for IHE faculty or staff is more likely to be related to specific research specialties and to take other forms, such as symposia, colloquia, and special opportunities working in others’ laboratories.

As previously noted, whether these 35 MSPs do in fact support such a large number of separate activities, or whether many of the initially separate activities are part of the same activity, is a topic for further investigation. One revelation that may occur is if multiple activities, within the same MSP, are found to be part of a strongly coordinated vision for an MSP as a whole. The situation might represent that of an MSP deliberately designing and implementing its various activities to move in the direction of reforming an entire K-20 system. Such ambitiousness also might deserve special attention because it could represent an important contribution by the MSP Program. On the basis of the available source material, the possibility of an MSP pursuing systems reform exists, in principle, in the case of 12 of the 35 MSPs that have reported four or more activities each (see Table 4). Ongoing evaluation work will monitor freshly collected evidence to test this supposition further.

### *What Are the Awardees Doing to Assess K-12 Student Achievement Outcomes?*

The early glimpse also covered the MSP awardees’ reports of whether and how they were assessing student achievement outcomes. Discussed earlier, this theme occupies a major place in the MSP Program, and the awardees had started to report this information not only in their annual reports but also as part of a specially designed management information system operated by the MSP Program.

This early information on student achievement also contributed to the initial design of the MSP-PE. Because of the time-consuming and difficult nature of assessing such outcomes, the evaluation team did not want to start on a redundant course. By the same token, the team was prepared to fill gaps and undertake its own analyses to provide the needed program assessment where needed. To inform these choices, the same 35 awards were characterized according to their own initial efforts in assessing student achievement outcomes.



TABLE 3  
Distribution of Activities Reported by Thirty-Five MSPs

Activities	No.	%	Illustrative Examples
I. The K-12 system			
IA. Work with K-12 students, classrooms, or curricula			
1. Support student enrichment activities	4		Science clubs to encourage HS students to enroll in math/science courses.
2. Implement new curricula, curriculum guides, or classroom technologies	6		S standards-based instructional materials for elementary schools.
	Subtotal		
	10	9.8	
IB. Work with K-12 teachers, administrators, or staff			
1. Provide inservice (prof devel) to existing K-12 classroom teachers	22		Lab-based prof devel for HS teachers, with IHE faculty spending time in HS classes.
2. Train teacher leaders, coaches, mentors, etc., to work with classroom teachers	7		Leadership action academies to train teacher leaders.
3. Implement cascading training system or learning community	4		Working with whole HS depts to change dept culture.
4. Train school administrators or staff	4		Administrators' institute to increase capability as instructional leaders.
	Subtotal		
	37	36.3	
IC. Work with K-12 policies and institutional structure			
1. Define and implement new standards, curriculum frameworks, or education policies	6		Comprehensive curriculum framework for partnering district.
2. Develop new assessment or other tools	2		New classroom assessments to accompany new curriculum guides.
	Subtotal		
	8	7.8	
II. The Undergraduate and Graduate (IHE)system			
IIA. Work with undergraduate and graduate students, classrooms, or courses			
1. Support student enrichment activities	8		Tuition stipends for students to enroll in MAT math and science courses.

(Continued on next page)

TABLE 3  
Distribution of Activities Reported by Thirty-Five MSPs (Continued)

Activities	Nq.	%	Illustrative Examples
2. Modify individual courses for existing undergraduates or graduates	5		Inquiry-based science incorporated into undergraduate and preservice courses.
3. Modify individual courses for existing K-12 teachers, administrators, or staff	3		Two-credit graduate course for school administrators.
	Subtotal		
	16	15.7	
IIB. Work with faculty, administrators, or staff			
1. Provide inservice (professional development) to existing IHE faculty	0		None
2. Train faculty leaders, coaches, mentors, etc., to work with IHE faculty	0		None
3. Implement cascading training system or learning community	0		None
4. Train IHE administrators or staff	0		None
	Subtotal	0.0	
IIC. Work with IHE policies and institutional structure			
1. Alter field of concentration or graduation requirements	6		New preservice course sequence.
2. Start or revise degree programs	5		New MSP master's program in curriculum and instruction.
3. Change IHE policies or encourage interorganizational collaboration	3		New language for tenure and promotion policies, enhancing faculty work with K
	Subtotal		
	14	13.7	
III. Families, community, and public and private institutions			
1. Organize family education or enrichment activities	6		Reform mathematics courses to engage adult learners.
2. Increase public awareness of mathematics and science education and its importance	1		Public awareness campaign to convey importance of math and science education.
	Subtotal		
	7	6.9	
IV. Interface between working with K-12 and IHE students, classrooms, or courses	3		Preservice course content aligned with HS math standards in partnering district.

(Continued on next page)

TABLE 3  
Distribution of Activities Reported by Thirty-Five MSPs (*Continued*)

Activities	No.	%	Illustrative Examples
V. Interface between working with K-12 and IHE faculty, administrators, and staff	0		None
VI. Interface between working with K-12 and IHE policies and institutional structure	7		Preservice program aims at PRAXIS exam and state math-science teaching licensure.
	Subtotal		
	10	9.8	
TOTAL	102	100.0	

*Note.* Source: MSPs' annual and evaluators' reports. MSP = Math and Science Partnership; HS = high school; IHE = institution of higher education; prof devel = professional development; dept. = department.

*The pattern of student achievement reporting, for 35 MSPs.* The tracking of the MSPs' efforts covered two dimensions: (a) whether and how an MSP was establishing any comparative framework for interpreting the achievement outcomes, and (b) the nature of an MSP's preliminary findings, if any—as interpreted and reported by the MSP itself. Regarding this latter point, most of the MSPs were reporting student achievement data that had occurred concurrently with their first or second year of work. Under this circumstance, some MSPs only directed their preliminary interpretations at defining baseline conditions. However, many of the MSPs went beyond this stage and tried to interpret emerging progress, also cautioning that different results might be expected following the full, five-year

TABLE 4  
Multiplicity of Activities Reported by Thirty-Five MSPs

No. of Activities	MSPs No.	%
None	3	9
Only 1 activity	6	17
2 activities	7	20
3 activities	7	20
4 activities or more	12	34
Total	35	100

*Note.* From MSPs' annual and evaluators' reports. MSP = Math and Science Partnership.

TABLE 5  
Concurrent Student Achievement Trends Reported by 35 MSPs

Framework for MSPs' Analyses	Percent Distribution, According to MSPs' Interpretation of Findings					Total
	No Analysis Yet	No Notable Differences in Trends Yet	Some Positive Findings	Some Negative Findings	Mixed Positive and Negative Findings	
None evident or unclear	19			—	—	19
MSP sites only (e.g., multiyear trends)	—	14	20	—	6	40
MSP compared to pre- established benchmark or district- or statewide groups	3	6	6	—	3	18
MSP and non-MSP groups compared (e.g., non-MSP classrooms, schools, or districts)	3	8	3	—	—	14
Stronger between- or or within-group designs		3	6	—	—	9
Total	26	31	35	—	9	100

*Note.* Source: MSPs' Annual and Evaluators' Reports.

period of most of the awards (among the 35 awards, 29 had five-year and six had four-year awards).

Collectively, the 35 IHEs' pattern of reports on student achievement fell into three broad groups, represented by the rows in Table 5:

1. Row 1, Table 5, shows that 7 of the 35 awardees had not yet reported data to address student achievement.
2. Row 2 shows that 12 of the 35 reported such data but not in any comparative context.
3. Rows 3 to 5 show that 16 of the 35 reported their MSP performance in the context of some non-MSP comparison.

Among the three groups, the MSPs that had not reported any data were still struggling with problems such as obtaining the needed data from state agencies or suffering from turnover in local evaluators.

Within the second group, the most common approach was to observe whether the MSP's own sites had changed over time, independent of referring to any comparative framework. As shown on Table 5, half of these MSPs interpreted the changes in a positive direction, because the scores had improved.



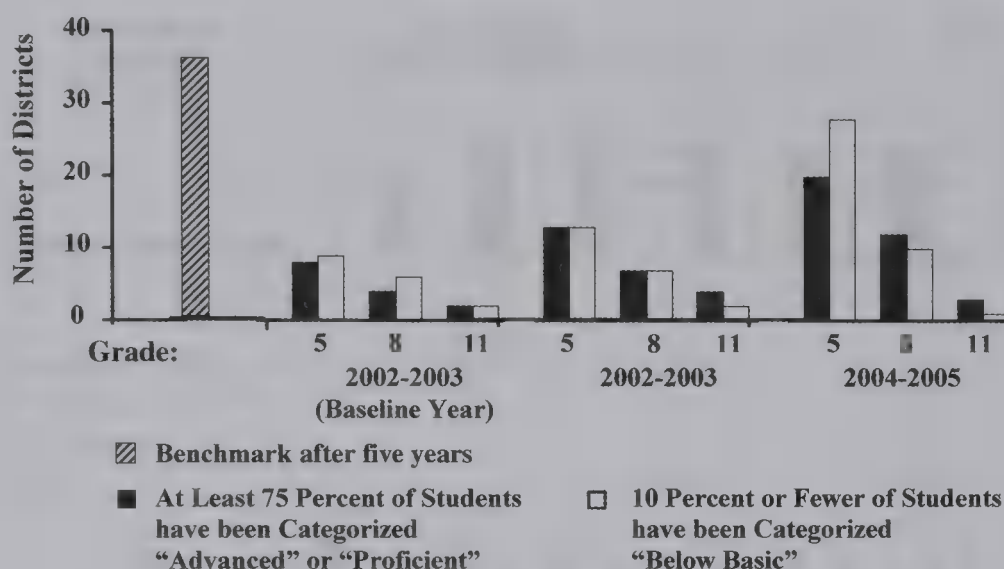


FIGURE 1 Number of participating districts meeting pre-established benchmarks in mathematics. *Note.* Source: MSPs' annual and evaluators' reports.

Within the third group, a common approach was for an MSP to establish a target or benchmark for the expected change and then to determine whether such a benchmark had been met. As one example, an MSP had 40 participating districts and had set an initial expectation that, by the end of the MSP's five-year period, 90% of these districts would exceed two benchmarks; that at least 75% of each district's students would have scored "advanced" or "proficient" on the state assessment in mathematics; and that 10% or fewer of the students would have scored "below basic." The MSP interpreted the trends for its first three years as suggesting that the districts were making progress toward these benchmarks in the 5th and 8th grades, but not in the 11th grade (see Figure 1).

The third group also included 3 MSPs that had stronger research designs to interpret the achievement results. One of these MSPs used a random-assignment design, with different classrooms assigned to "treatment" or "no-treatment" conditions. The "treatment" provided teachers with inservice training. The findings showed no differences in the mathematics scores between the two groups (though they differed significantly on the overall state assessment—see Figure 2a), but the evaluation also showed no significant differences in the instruction provided to these two groups, either (see Figure 2b). The MSP therefore concluded that it had to do more work in making the "treatment" more potent before differences in achievement could be expected.

Another of the 3 MSPs used a cross-sectional design that nevertheless could support a more robust interpretation. For this MSP, the achievement data came from a science test given to all fifth-grade students in all of a partnering district's

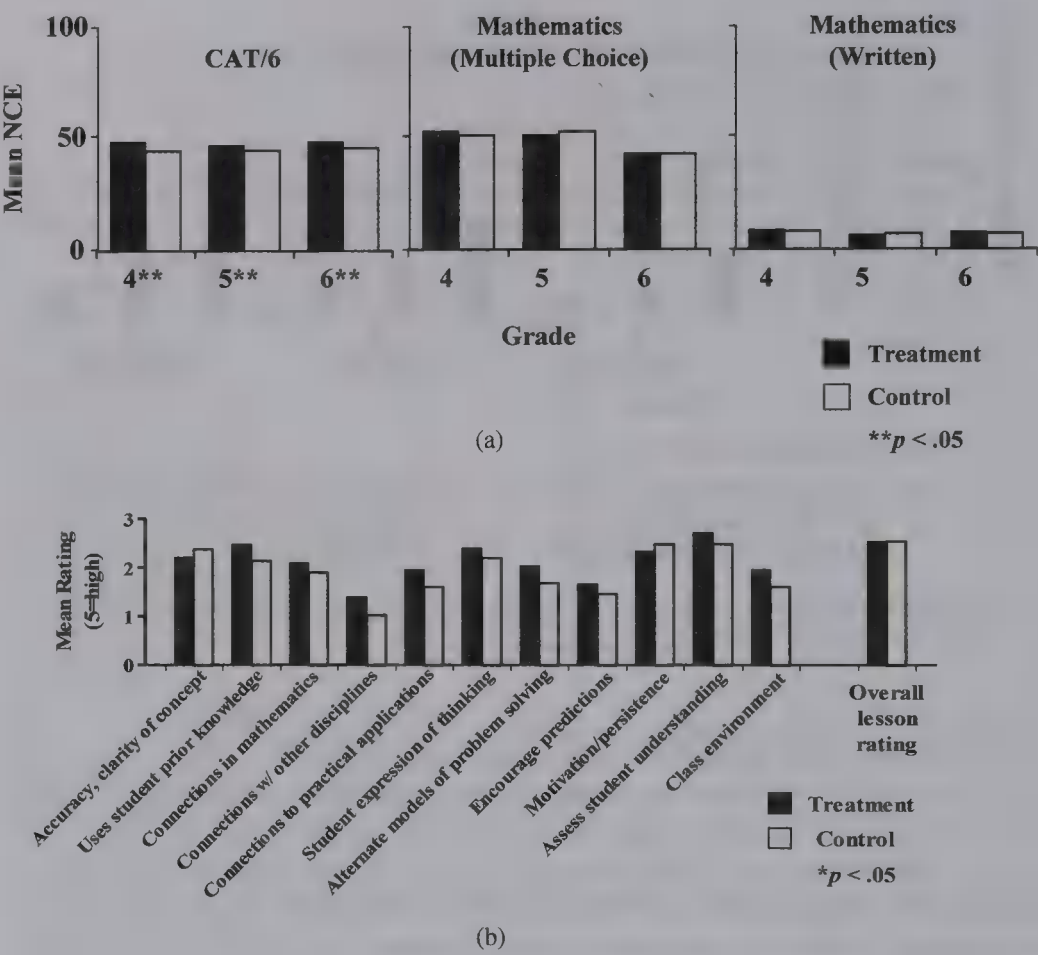


FIGURE 2 A Math and Science Partnership (MSP) that randomly assigned classrooms to “treatment” and “nontreatment” conditions. *Note.* The “treatment” provided teachers with inservice training in mathematics (only). The findings showed no differences in the mathematics scores between the two groups (though they differed significantly on the overall CAT/6 state assessment), but the evaluation also showed no significant differences in the instruction provided to these two groups, either. The MSP therefore concluded that it had to do more work in making the “treatment” more potent before differences in mathematics achievement could be expected. (a) Students’ math scores from two randomly assigned groups of schools (total number of students was not reported). (b) Mathematics classroom observations from two randomly assigned groups of schools, spring 2005 ( $n = 35$  classrooms). Source: MSPs’ annual and evaluators’ reports.

schools. The test contained four separately scored strands, and for two of the schools in this district, the MSP had helped the classrooms to strengthen their science instruction but only on a limited number of science topics. For school “L NES” (see Table 6), the topics coincided with Strands 2 and 4 of the science test, whereas for school “SHAR,” the topics coincided with Strands 1 and 3.

TABLE 6  
A Helpful Within-Group Design

School	% of Students Achieving Proficiency in Grade 5 Science December 2005			
	Strand 1	Strand 2	Strand 3	Strand 4
CEHE	59.7	49.0	42.5	56.4
CENT	48.3	51.8	36.0	53.3
COSP	54.6	49.1	54.9	54.7
EAST	55.5	46.7	42.2	52.3
EBEN	52.0	48.1	44.5	53.1
HARM	64.4	54.0	43.8	62.9
LNES	60.6	60.5	51.6	<b>67.5</b>
LAES	61.5	54.1	57.9	62.3
MONT	53.3	46.7	41.7	46.0
TMO	72.4	<b>60.8</b>	48.9	65.9
SCOT	58.6	54.5	39.5	46.0
SHAR	<b>71.1</b>	51.0	<b>56.9</b>	54.7
SHEP	62.5	58.4	54.8	57.7
TCES	51.7	46.6	40.1	50.1
TRES	56.8	53.4	44.8	56.7
UGES	51.1	57.9	56.7	57.1
WHES	57.6	58.4	53.8	65.2
DISTRICT	<b>58.2</b>	<b>52.9</b>	<b>47.5</b>	<b>57.2</b>

*Note.* For this Math and Science Partnership (MSP), the achievement data came from a science test given to all fifth-grade students in all of a partnering district's schools. The test contained four separately scored strands, and for two of the schools in this district, the MSP had helped the classrooms to strengthen their science instruction but only on a limited number of science topics. For school "LNES," the topics coincided with Strands 2 and 4 of the science test, whereas for school "SHAR," the topics coincided with Strands 1 and 3.

The results showed that the fifth graders in these two schools performed better on the alternating strands that coincided with the strengthened instruction than on the other strands that did not coincide. The higher scores placed the two schools' performance at or near the top of all of the other schools in the district, on the same coinciding strands, whereas the lower scores placed the two schools near the average of the other schools. Source: MSP's annual and evaluator's reports.

The results showed that the fifth graders in these two schools performed better on the alternating strands that coincided with the strengthened instruction than on the other strands that did not coincide. The higher scores also placed the two schools' performance at or near the top of all of the other schools in the district, on the same coinciding strands, whereas the lower scores place the two schools near the average of the other schools.

*Comments on MSPs' existing assessments of student achievement.* The available secondary reports suggest that the 35 MSPs are making progress in collecting and reporting student achievement data. At the same time, the MSPs

have not necessarily fully confronted the challenges in analyzing these data. Some of the challenges (and potential remedies) are as follows.

First, some of the MSPs have reported difficulties in establishing trends over time because of changes in the state assessment instrument. These MSPs, and especially their partnering districts, could more closely collaborate with their respective state education agencies, to understand how the states themselves may be calibrating the scores from their different tests over time. Most states may be doing such calibration, in light of the requirements under *No Child Left Behind*.

Second, many of the MSPs have reported districtwide data, although the MSPs may not have implemented their activities on a districtwide basis. A similar situation can exist at the school level, where the MSP reports may have reported aggregate school-level data, even though the MSP's activities only have involved some of the classroom teachers. In either situation, scale-up may still be occurring, but until fully scaled, the MSPs may need to match more closely their achievement data with the venues in which the MSP activities have taken place.

Third, those MSPs that have chosen to define pre-established benchmarks for later comparison to actual performance have not usually discussed any rationale for selecting their particular numeric benchmark. For instance, the MSPs do not discuss whether such benchmarks as "improving performance by 5% each year" might be too conservative or overly ambitious. Where benchmarks are to be used, some discussion and rationale for the cut-points selected would be helpful.

Fourth, many MSPs report scores for multiple grade levels and for both science and mathematics assessments. MSPs in this situation might want to consider setting another type of benchmark: whether all scores are expected to improve or whether only one or a few are.

Finally, most of the research designs make it difficult for an MSP even to begin the needed attribution process. Because of the large size of the MSPs and the number of students and teachers being affected, possibly the MSPs could implement some small-scale research, focusing on a few classrooms or schools, that would nevertheless use more robust experimental designs.

### *What Is the Framework and Design for the MSP-PE?*

These early glimpses provided one starting point for designing the MSP-PE. A second starting point was a set of six questions and topics that the evaluation is to address:<sup>6</sup>

---

<sup>6</sup>The six questions and topics appear in the original solicitation for the MSP-PE award and define its goals.



*Questions:*

1. How has the MSP Program affected, influenced, or been associated with changes in the K-12 mathematics and science teaching force, K-12 student achievement in math and science, and other outcomes associated with the program?
2. What factors or attributes appear to have accelerated or constrained progress in the MSP Program's achievements?
3. How have disciplinary faculty (math, science, and engineering) from IHEs participated in the MSP Program, and what has been their role in the program's achievements?

*Topics:*

4. The MSP's Program features, including MSP-related discoveries and innovations in math and science education worth developing on a large scale.
5. The processes influencing, interfering, or associated with the outcomes of the features.
6. The conditions associated with the demonstrated quality and innovativeness of the MSP Program.

The desired evaluation design needs to address these questions and topics, also accounting, as previously discussed, for the diversity of the MSPs' ongoing activities and the MSPs' efforts to assess student achievement outcomes. The design also needs to suit the MSP-PE as a *program evaluation*, representing the workings of the MSP Program as a whole, not simply the sum of individually funded awards, and certainly not evaluating any of the awards individually.<sup>7</sup>

Toward this end, the design of the evaluation has involved two stages. First, the evaluation team has formulated a policy-oriented framework, covering K-20 education in mathematics and science by identifying the institutions that play a role in educating the science and engineering workforce, including producing newly qualified K-12 mathematics and science teachers as well as assisting existing teachers. Second, rather than forcing itself into the confines of a single evaluation study, the team has begun implementing its evaluation as a series of substudies. Progress on both stages is described next.

*A preliminary evaluation framework: Pathways through a multi-institutional, K-20 world.* First and foremost, the nature of the MSP Program points

---

<sup>7</sup>Each MSP has an ongoing "project" evaluation to serve this purpose. To date, most of the project evaluators are collecting data and providing "formative" feedback to their host MSP. Some but not all of the evaluators eventually plan to conduct "summative" evaluations. Regardless, the project evaluations will not address cross-project findings or lessons, which are the main thrust of the MSP-PE.

to the need for a multi-institutional framework. At minimum, the framework should include a partnership among the institutions involved in a K-20 span of mathematics and science education (see Figure 3a).

The emphasis on K-12 student achievement, as well as the mixture of activities sanctioned by the MSP Program, then suggest how the actions of these institutions might be related through a *series of pathways*. Existing teachers and faculty in the partnering K-12 and IHE institutions are part of these pathways (see Figure 3b). Students traverse through these pathways, at first completing their K-12 education. Some may enter the job market after high school, but others will proceed to undergraduate and graduate careers within IHEs, followed by employment opportunities that include becoming K-12 teachers. Those K-12 teachers will in turn teach a new generation of K-12 students. Other graduates will pursue careers as scientists, engineers, computer specialists, and other professions in the S&E workforce.

The pathways and multi-institutional scope suggest a systems framework within which successful student achievement through the K-12 grades is essential for the pathways to work—hence the MSP Program's focus on improving such achievement. However, other transitions, such as successful entrance into undergraduate and graduate programs, as well as successful graduation from these programs and entrance into workforce positions, also are important. The evolving framework therefore points not only to the institutions involving K-20 education but also the pathways that represent the interactions among these institutions and the multi-institutional transitions that people must negotiate successfully.

*How the preliminary framework serves the needs of program evaluation.* This preliminary framework provides a unifying scope for evaluating the MSP Program.

First, although the framework has been set forth at a broad level, much detail can be added readily. These details, covering specific activities within an institution or the collaboration between institutions, can enable the evaluation to test hypotheses about lessons learned in response to the six evaluation questions and topics previously enumerated. Appropriately reflecting the level of concern raised by these six questions and topics, the framework enables the evaluation to be directed at the MSP Program as a whole.

Second, although the MSP awards may comprise a heterogeneous portfolio, the possibility is that all of the awards in the portfolio may be conceptually locatable within the framework, thereby bringing unity to the portfolio, despite its diversity. If successful in capturing the entire portfolio within a single framework, the MSP-PE can more readily derive conclusions about the MSP Program as a whole—for example, whether the MSPs' activities have fallen within particular portions of the pathways but left other portions uncovered.

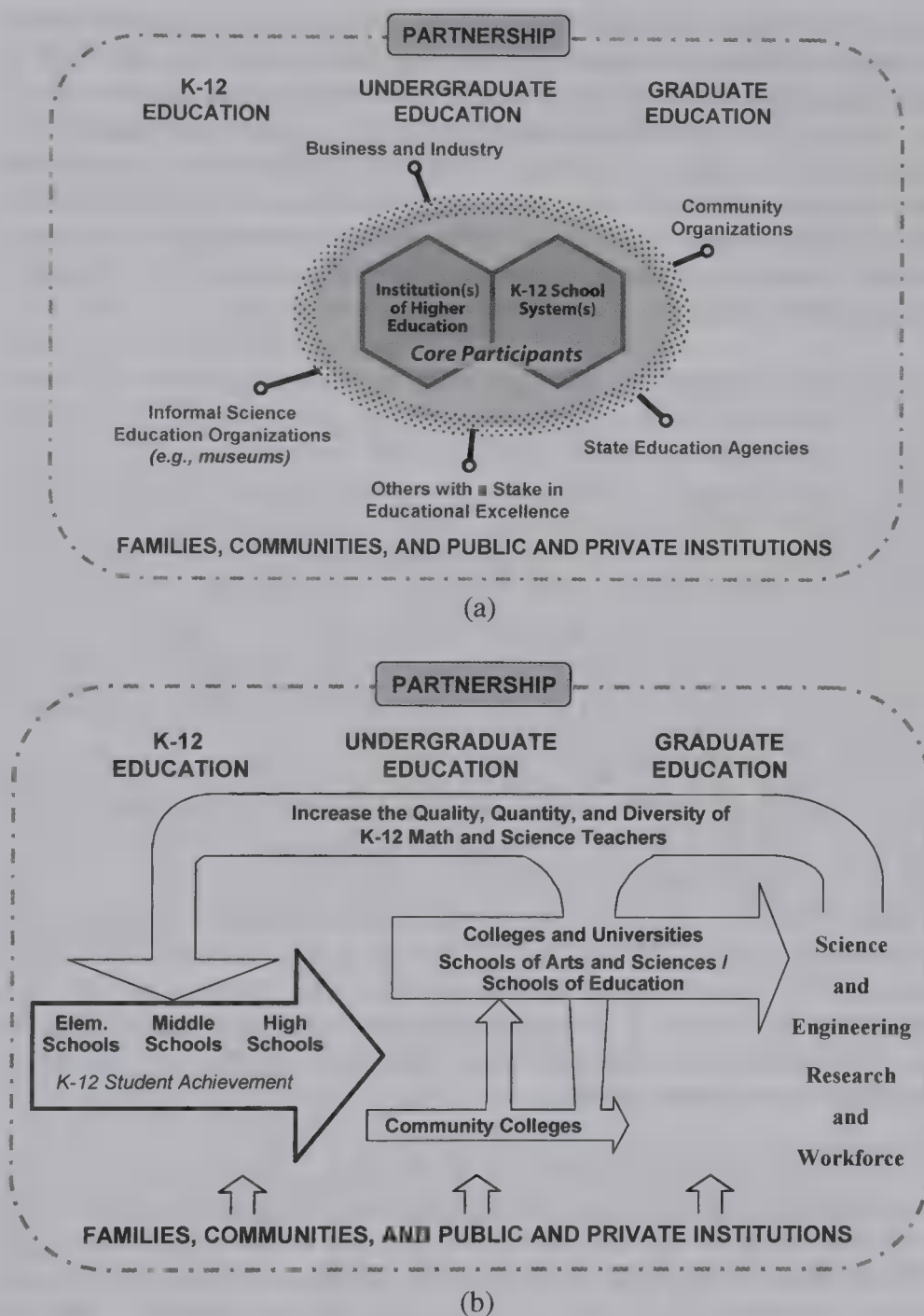


FIGURE 3 A preliminary framework for the Math and Science Partnership program evaluation.

Third, the framework still points to the critical role of assessing K-12 student achievement. Given the disparate efforts among the individual MSPs, the MSP-PE needs to give high priority to such an assessment, with a program-wide perspective and not just reporting results from individual MSPs. In particular, the MSP-PE's assessment also needs to include some type of comparative design. Two basic designs quickly surface. One would contrast the performance of MSP and non-MSP entities (see Wong, & Socha, 2008). A second could test a cross-MSP correlation between the range of intensities of MSP activities with the range of student achievement outcomes (see Dimitrov, 2008/this issue).

*A series of substudies.* More generally, the framework can provide a unifying scope for an entire series of substudies, all undertaken as part of the MSP-PE. A substudy strategy has appeal because of both the heterogeneity of the MSP Program's portfolio and the R&D nature of the MSP Program. The two conditions mitigate against the more conventional, singular evaluation study. For instance, each substudy can have its own design, customized to focus on a different but essential part of the framework, with (quantitative and qualitative) meta-analytic strategies serving as the main methods for amassing and synthesizing evidence about the MSP Program's portfolio. The meta-analytic strategies are to treat each of the MSP awards as studies or sets of studies (depending on the multiplicity and separateness of an MSP's activities) in and of themselves.

The substudies also will have their own phases, evolving over time. The first phase of a substudy may be limited to a review of the literature and the analysis of secondary materials about the MSP Program and its awardees. A later phase can incorporate original field data collected by the MSP-PE itself.

Some of the MSP-PE's substudies are contained in this issue. The study on MSP Partnerships (Scherer, 2008/this issue) is limited at this juncture to information contained in the awardees' annual reports. Somewhat different but still in its early phase, the study on K-12 student achievement trends (Dimitrov, 2008/this issue) was based on the initial data (Wave I and Wave II) collected by the MSP's management information system.

*Why the evaluation framework is still evolving.* In like manner, the evaluation framework also is evolving. Among other topics, the framework does not yet try to distinguish among mathematics, science, and engineering education. The MSPs, however, vary in their coverage of these fields, with about 40% focusing on both mathematics and science education but 40% on mathematics only and 20% on science only. "Science," of course, also consists of multiple disciplines with different educational features. The potential differences in the institutional structure and pathways among these fields raise the possibility of needing a more refined evaluation framework. How this and other challenges are to be confronted is part of the MSP-PE's ongoing work.



*Summary of early progress by the MSP-PE.* This study has described the MSP Program and the initial stages of its program evaluation. The study has discussed four important themes underlying the overall program and has presented a preliminary characterization of 35 MSPs supported as part of the program's first two cohorts of awards. The study has then turned to the challenge of developing an evaluation design, which has followed two stages: an evaluation framework highlighting interinstitutional relationships in a K-20 system of mathematics and science education, and a series of substudies, each focusing on a slightly different part of the framework. Some of the early substudies appear as studies in this special issue (also see the introduction to this special issue).

## ACKNOWLEDGMENT

This study is one in a series of substudies for the Math and Science Partnership Program Evaluation (MSP-PE) conducted for the National Science Foundation's MSP Program. The MSP-PE is conducted under Contract No. EHR-0456995. Since 2007, Bernice Anderson, Ed.D., Senior Advisor for Evaluation, Directorate for Education and Human Resources, has served as the National Science Foundation Program Officer. The author, from COSMOS Corporation, is Robert Yin.

The MSP-PE is led by COSMOS Corporation in current partnership with George Mason University (GMU) and Brown University. Robert K. Yin (COSMOS) serves as Principal Investigator and Jennifer Scherer (COSMOS) serves as one of three Co-Principal Investigators. Additional Co-Principal Investigators and their collaborating institutions (including discipline departments and math centers) are Patricia Moyer-Packenham (USU, formerly GMU) and Kenneth Wong (Brown).

## REFERENCES

- Dimitrov, D. (2008/this issue). Initial trends in MSP-related changes in student achievement with MIS data. *Peabody Journal of Education*, 83(4), 637–653.
- Domestic Policy Council, Office of Science and Technology Policy. (2006, February). *American competitiveness initiative: Leading the world in innovation*. Washington, DC: Author.
- National Academy of Sciences, Committee on Science, Engineering, and Public Policy. (2007). *Committee on prospering in the global economy of the 21st century* (Report). Washington, DC: Author.
- National Science Board, *2020 Vision for the National Science Foundation*, Arlington, VA, December 28, 2005, NSB-05-142.
- National Science Foundation. (2001). *Math and Science Partnership (MSP) Program* (Program solicitation, NSF-02-061). Arlington, VA: Author.
- National Science Foundation. (2002). *Math and Science Partnership (MSP) Program* (Program solicitation, NSF-02-190). Arlington, VA: Author.
- National Science Foundation. (2003a). *Math and Science Partnership (MSP) Program*, (Program solicitation, NSF-03-605). Arlington, VA: Author.

- National Science Foundation. (2003b). *Math and Science Partnership* (Program solicitation, NSF-03-541). Arlington, VA: Author.
- National Science Foundation. (2006). *Math and Science Partnership (MSP) Program*, (Program solicitation, NSF-06-539). Arlington, VA: Author.
- P.L. 107-368, National Science Foundation Authorization Act of 2002, December 19, 2002 (§9(3)).
- Scherer, J. (2008/this issue). A review of instruments to evaluate partnerships in math and science. *Peabody Journal of Education*, 83(4), 611–636.
- Sroufe, G. (2006, April/May). [OIA Info Memo]. Washington, DC: American Educational Research Association.
- U.S. General Accountability Office. (2006, May 3). *Science, Technology, Engineering, and Mathematics Trends and the Role of Federal Programs* (Testimony before the Committee on Education and the Workforce, House of Representatives, GAO-06-702T).
- U.S. House of Representatives, Committee on Science, Subcommittee on Research. (2003, October 30). Implementation of the Math and Science Partnership Program: Views from the field. *Hearings*, 107th Congress, 2nd Session.
- Wong, K., & Socha, T. (2008/this issue). A pilot study to identify comparison schools for MSP participating schools: Preliminary findings on one math/science partnership. *Peabody Journal of Education*, 83(4), 654–673.
- Yin, R., Hackett, E., & Chubin, D. (2008/this issue). Discovering “what’s innovative”: The challenge of evaluating education R&D efforts. *Peabody Journal of Education*, 83(4), 674–690.

# A Review of the Literature on Mathematics and Science Teacher Quality

Johnna J. Bolyard  
*West Virginia University*

Patricia S. Moyer-Packenham  
*Utah State University*

A large body of literature exists that examines teacher quality characteristics and the relationship of indicators of those characteristics to teacher effectiveness. This existing research literature broadly views teacher quality research without illuminating specific areas of teacher quality, such as mathematics and science. In an effort to focus the literature base for researchers and policymakers more narrowly, this review specifically examines teacher quality as it relates to mathematics and science teaching and learning. The review highlights key policy and practitioner perspectives, provides a focused synthesis on current research findings on mathematics and science teacher quality, and suggests areas of research that are limited in the literature.

Recently, K-12 education has been engaged in a struggle to staff schools with qualified teachers, particularly in areas such as mathematics and science. This growing struggle has prompted concerns about teacher supply, as evidenced in the landmark report published by the National Commission on Teaching and America's Future (1996), and high-profile examinations of and policy statements on the status of teacher quality (Mitchell, Robinson, Plake, & Knowles, 2001; U.S. Department of Education, 2002). These examinations are not surprising when a large body of empirical research has identified differences in the quality of the teacher as explaining more of the variation in student achievement than

any other school-based factor (Goldhaber & Brewer, 1997b; Sanders & Rivers, 1996; Strauss & Sawyer, 1986). This recognition has led to major efforts funded by the federal government, under the direction of agencies such as the National Science Foundation and the Department of Education, to initiate *Math and Science Partnership Programs* for the purpose of improving mathematics and science teaching and learning in K-12 schools.

Researchers, policymakers, and educators have historically viewed teacher quality from differing perspectives. For example, from a *researcher's* point of view, teacher quality is operationalized as a construct and variables are identified and examined in relation to outcome measures. For a *policymaker*, teacher quality provides a benchmark against which individuals can be identified as meeting or not meeting a given standard of quality (Blank & Langeson, 1999). School administrators view teacher quality as a means of finding the right educator (i.e., the one with the most potential, based on a set of qualities and skills) for the job. For *educators* in different positions within the educational system, teacher quality takes on different meanings. For the classroom teacher, teacher quality may be viewed as a continuous process of self-renewal and professional development where one works to impact and improve the quality of one's own teaching. A teacher educator may view a quality teacher as one who has a strong foundational knowledge of content and pedagogy that can be built upon and strengthened throughout his or her career.

With these perspectives in mind, it is easy to see how different views have emerged within the construct of teacher quality. Yet within these perspectives are overlapping themes which indicate that, perhaps, what appears on the surface to be a divergence of views actually masks a lack of clarity about what is meant by and known about teacher quality. Because of these differing perspectives, researchers, policymakers, and educators draw on different literature to make decisions about mathematics and science research, policy statements, and educational initiatives and interventions. Although other authors have reviewed the body of literature that identifies and examines variables believed to be indicators of teacher quality and the relationship of these variables to teacher effectiveness (Rice, 2003; Wayne & Youngs, 2003), this existing literature broadly reviews teacher quality research without specific emphasis on any subject area. In an effort to bring together and focus these different literatures, this article specifically examines teacher quality as it relates to mathematics and science teaching and student outcomes. Our purpose was to create a document that could be used by researchers, policymakers, and educators as a summary of current findings on mathematics and science teacher quality.

In the sections that follow, we outline our methodology for selecting documents for inclusion, provide a synthesis on mathematics and science teacher quality from these documents, and summarize key findings from the research. The final section discusses general implications and suggests areas for further research. One



important item to note is that the scope of this review did not seek to encompass all variables, actions, influences, and conditions of mathematics and science teacher quality. The primary goal of our article was to focus on individual characteristics of teachers. Therefore, the selection of literature distinguished between those studies that focused on characteristics of individual mathematics and science teachers and those that focused on characteristics of the teacher population. For example, research on characteristics of the teacher population that may influence teacher quality, such as the recruitment of a diverse teaching force and the supply and demand of the mathematics and science teacher population, was not part of this review. Although these broader issues are important, an examination of population characteristics of teachers and teacher quantity was beyond the scope of this article. This review provides a systematic and focused examination of the teacher quality literature as it relates to characteristics of mathematics and science teachers and student outcomes.

## METHODS

We began our review with an exhaustive search of electronic databases. This search led us to several meta-analyses and reviews including those conducted by Rice (2003) and Wayne and Youngs (2003) on teacher characteristics and attributes; Cochran-Smith and Zeichner (2005); Wilson and Floden (2003); and Wilson, Floden, and Ferrini-Mundy (2001) on teacher preparation; and Greenwald, Hedges, and Laine (1996) on school resources. These large-scale analyses provided a foundation for further investigation into the literature. From this initial search, we began a more extensive review using basic search procedures and following standard criteria for a comprehensive literature search (Boote & Beile, 2005). This process included library searches (both electronic and manual) in educational databases such as ERIC, PsycInfo, and Social Sciences Index using search terms including *teacher*, *teacher characteristics*, *teacher quality*, and *teacher effectiveness*. We searched online mathematics and science organizations for statements and position papers on teacher quality specific to these two disciplinary foci.

As part of the process for inclusion in this article, we were selective in including primary documents and widely available information from Web-based and library sources. Documents included from electronic formats were obtained from the main Web sites of the major agencies and organizations with the authority to speak for the organization. For example, we used the Web sites of the U.S. Department of Education, the National Science Teachers Association (NSTA), and the National Council of Teacher of Mathematics (NCTM) as sources of information on the positions of these organizations, rather than obtaining information

from other sources that quoted, reported, or described the positions of these organizations.

In our selection of library and research publications and documents, we included works that were empirical, meta-analyses, and literature reviews that appeared as peer-reviewed documents. These resources were selected based on the methodological rigor of the studies, the frequency with which studies were cited and referenced by other researchers, and the focus of all or part of the research on characteristics of teacher quality. From these studies of general teacher quality, we further examined the studies and documents for information that identified teacher quality characteristics related to mathematics and science teachers. Because of the large number of documents uncovered, we further limited this review to research that targeted characteristics of teachers of mathematics and science, and used measures of students (i.e., mathematics and science achievement or student attitudes) as outcome variables. We chose student outcomes as the dependent variable because improving students' learning and educational experiences is a common goal among educational stakeholders. In total, we reviewed approximately 150 documents.

## FINDINGS

The following sections present the findings of our review. We have organized the findings into two major sections: (a) key policy, public, and practitioner documents focused on mathematics and/or science teacher quality, and (b) relevant studies that correspond to mathematics and/or science teacher quality characteristics. In the first section, key policy, public, and practitioner documents, such as No Child Left Behind (NCLB) legislation, National Board Certification Standards, NCTM, and NSTA descriptions of teacher quality, are highlighted and discussed. In the next section, we present a review of the research on mathematics and science teacher quality using six individual teacher quality characteristics: general ability; experience; pedagogical knowledge; subject matter knowledge; certification status; and teacher behaviors, practices, and beliefs. These categories of individual teacher quality were chosen based on their frequent use in large-scale meta-analyses and reviews of the literature on general teacher quality (Cochran-Smith & Zeichner, 2005; Darling-Hammond, 2000; Rice, 2003; Wayne & Youngs, 2003; Wilson & Floden, 2003; Wilson et al., 2001). Although other factors, such as professional development, have also been examined in relation to teacher quality, professional development was defined by the authors as a *condition* or *intervention* used to impact teacher quality characteristics and was beyond the scope of the current review.

## Key Policy, Public, and Practitioner Documents on Mathematics and Science Teacher Quality

Teacher quality, and specifically mathematics and science teacher quality, has been the focus of much public debate in education. Federal documents such as NCLB and related educational policies have escalated the focus on the quality and quantity of mathematics and science teachers in the United States. In response to this legislation and the surrounding discussion, several professional organizations have produced position statements outlining their view of a quality mathematics and science teacher.

### Federal Policy and Report Documents

The NCLB Teacher Quality mandate states that a *highly qualified teacher* holds at least a bachelor's degree; holds full certification or has passed a teacher licensing examination (as dictated by a state licensing agency); and holds a license to teach that is not classified as emergency, temporary, or provisional. Further, "highly qualified" teachers must demonstrate competence in subject knowledge and teaching skills. Elementary teachers new to the profession must demonstrate competence in a subject such as mathematics, reading, writing, and other areas of the curriculum by passing a "rigorous State test" (U.S. Department of Education, 2002, p. 5). New middle or secondary teachers must demonstrate competency in all of the subjects they teach by passing a state subject test or completing a degree in the subject, coursework equivalent to a degree in the subject, or advanced certification or credentialing (U.S. Department of Education, 2002). Existing teachers must demonstrate competence either through an examination or based on a "high objective uniform State standard of evaluation" (U.S. Department of Education, 2002, p. 5). Although this portion of the general NCLB mandate does not specifically address mathematics and science teacher quality directly, its requirement of demonstrated competency in a subject area sends a clear edict to mathematics and science educators that content knowledge does matter.

Other federal documents echo NCLB's focus on the importance of quality teaching, particularly in the areas of mathematics and science. In July 1999, U.S. Secretary of Education Richard Riley announced the appointment of the National Commission on Mathematics and Science Teaching for the 21st Century (the Glenn Commission) to investigate the quality of mathematics and science teaching in the country and to examine ways to increase the number and quality of mathematics and science teachers in K-12 schools. The resulting report, *Before It's Too Late* (National Commission on Mathematics and Science Teaching for the 21st Century, 2000), highlighted the importance of quality mathematics and science education in preparing students to be competitive in an increasingly global society. The report identified improvement of teaching as the best way to achieve that goal and

described a vision of high-quality teaching that places deep content knowledge at its foundation. More recently, the 2006 American Competitiveness Initiative established the National Math Panel to bring together experts in mathematics, education, and cognitive science to make recommendations on the most effective methods for teaching mathematics (Office of the U.S. Press Secretary, 2006). The American Competitiveness Initiative calls for “Math Now” programs for the purposes of translating and disseminating the findings of the National Math Panel to classroom teachers.

### National Board for Professional Teaching Standards

Another perspective on mathematics and science teacher quality is that identified by the National Board for Professional Teaching Standards as *accomplished teaching*. Accomplished teaching is based on a set of standards for each area of teaching and five core propositions which include that teachers are committed to their students’ learning, know the content of their subjects as well as how to teach it, carefully monitor students’ learning, reflect on their practices, and are members of professional learning communities (National Board for Professional Teaching Standards, 2002). For teachers of mathematics and science, there are four certificate and standards areas categorized by students’ age and teachers’ subject: Generalist Early Childhood (ages 3–8; includes mathematics and science), Generalist Middle Childhood Certificate (ages 7–12; includes mathematics and science), Mathematics or Science Early Adolescence Certificate (ages 11–15), and Mathematics or Science Adolescence and Young Adulthood Certificate (ages 14–18+). To demonstrate accomplished teaching, teachers complete timed subject-area exams and create a teaching portfolio that contains videotapes of their teaching, evidence of student learning products, and a detailed analysis of their teaching practices. A surface examination of the requirements for *highly qualified teachers* (NCLB) compared with *accomplished teachers* (National Board for Professional Teaching Standards) appears to indicate that, in addition to a bachelor’s degree, full state certification, and competency in the subject area, an accomplished teacher must document and analyze student learning and teaching practices.

### Professional Organizations

Following the NCLB mandate, a variety of position papers and statements emerged from professional organizations and societies to clarify what it means to be a “highly qualified teacher of mathematics or science.” These documents describe what knowledge should be demonstrated by a quality mathematics and science teacher. For example, in 1991, the Mathematical Association of America released



*A Call for Change: Recommendations for the Mathematical Preparation of Teachers of Mathematics*, which called for a change in the mathematical preparation of prospective teachers. The report lists standards in seven content areas and recognizes that preparation must also include mathematical pedagogy (Leitzel, 1991). The Conference Board of Mathematical Sciences released in 2001 *The Mathematical Education of Teachers*. The report, described as an augmentation of the Mathematical Association of American report, focuses on two major themes: the substance of school mathematics and the nature of mathematical knowledge needed by teachers (Conference Board of Mathematical Sciences, 2001). The report recommends that mathematics coursework for prospective teachers deepen their knowledge of the mathematics they teach, focus on a coherent development of mathematical ideas, and develop the habits of mind of mathematical thinking. Further, the report recommends specific quantities of mathematics coursework for teachers at the elementary (at least 9 semester hr), middle (at least 21 semester hr), and high school levels (the equivalent of an undergraduate major), specifying that this coursework should be relevant to the mathematics that teachers will teach. Other recommendations include making teacher education an important part of the work of mathematics departments and fostering cooperation between mathematics and mathematics education faculty, two- and four-year institutions, and higher education and K-12.

The *National Science Education Standards* (National Research Council, 1996) describe quality science teachers as those who create supportive, active learning environments for their students, use assessments to inform and guide their teaching, participate in professional learning communities, and are committed to lifelong learning of science and science teaching and learning. The Council of Scientific Society Presidents described a “well-qualified” mathematics or science teacher as someone who understands mathematics and science deeply, uses instructional techniques that facilitate students’ mathematical problem solving and communication, and commits to lifelong learning and improvement of his or her practice (Council of Scientific Society Presidents, 2004). These recommendations show that teacher quality includes teacher education courses as well as courses in the discipline.

Leading mathematics and science teacher education associations and organizations, including the NCTM, the NSTA, the Association of Mathematics Teacher Educators, and the Association for Science Teacher Education have similarly addressed issues of teacher quality in mathematics and science. NCTM (1991) outlines standards for mathematics teaching in *Professional Standards for Teaching Mathematics*. These standards are based on a framework for teaching that highlights the important decisions teachers make in their work including the selection of mathematical tasks designed to facilitate significant learning, establishing effective classroom discourse around mathematics, creating inviting and safe learning environments, and making informed decisions about future instructional

goals. The *Professional Standards*' vision of effective professional development and training for teachers includes experiences that model good mathematics teaching, develop knowledge about mathematics, mathematics pedagogy, and students, and facilitate the continued development of teachers' practice. In a more recent position statement, titled "Highly Qualified Teachers," NCTM further outlines the qualifications required of a high school, middle school, and elementary school teacher of mathematics. These include the completion of coursework equivalent to a major in mathematics for high school teachers, coursework equivalent to at least a minor in mathematics for middle school, and at least the equivalent of three college-level mathematics courses for elementary and all other teachers of mathematics (NCTM, 2005, ¶2). An NSTA (2004) position statement titled "Science Teacher Preparation" states that all teachers entering the profession need a deep understanding of pure and applied science and the knowledge necessary to teach it meaningfully (¶2). In their position statement on science teacher preparation, the Association for Science Teacher Education (2004) describe quality science teachers as those who have deep understanding of science content, its applications, and history, as well how students learn science concepts and develop skills and dispositions necessary to engage in scientific inquiry.

Like the statements of other governmental and professional organizations, the statements of these organizations emphasize the importance of content knowledge. In addition, they emphasize the need for well-qualified teachers to develop an understanding of the subject appropriate to the level at which they teach as well as an understanding of how to effectively use that knowledge to create opportunities for learning.

## RESEARCH ON MATHEMATICS AND SCIENCE TEACHER QUALITY

Perspectives on mathematics and science teacher quality can be seen in variables used by researchers to operationalize the teacher quality construct. This section discusses the results of an extensive review of the literature focusing on characteristics researchers have used to operationalize individual teacher quality and measures used to examine relationships to teacher effectiveness. From this review, we have identified six primary characteristics studied frequently as indicators of individual teacher quality including: general ability; experience; pedagogical knowledge; subject knowledge; certification status; and teacher behaviors, practices, and beliefs.

### General Ability

Many studies examining the relationship between teachers' effectiveness and their academic and verbal ability use large-scale data sets that do not distinguish among

teachers of specific subjects. Several of these studies have generally found a positive relationship between teachers' academic ability and student achievement (Greenwald et al., 1966; Hanushek, 1971; Strauss & Sawyer, 1986). Of these studies, a few include students' mathematics achievement either as part of a composite achievement score or as a separate achievement indicator. For example, Ehrenberg and Brewer (1994, 1995) found that having a teacher who had attended a more selective undergraduate institution was statistically significantly associated with higher gains in the average of high school students' test scores in several areas, including mathematics. However, the relationship between teachers' verbal ability and student achievement varied depending on the teachers' race and the students' race and grade level (Ehrenberg & Brewer, 1995). Ferguson (1991) found that teachers' performance on a test measuring basic literacy skills explained one fifth to one fourth of the variation in 1st-, 3rd-, 5th-, 7th-, 9th-, and 11th-grade students' reading and mathematics achievement scores across 900 school districts; teachers' test scores were "the most important school input for both math and reading" (p. 475). Ferguson and Ladd (1996) found a positive relationship between average teacher ACT scores and 4th-grade student achievement in reading and mathematics. Although the effects were positive in both areas, the result for reading was statistically significant.

Studies of the relationship between teachers' general and verbal ability and student achievement often do not specifically focus on mathematics and science teacher quality. However, a few studies do use student mathematics performance as an outcome measure and these studies generally point to evidence of a positive relationship between teachers' general and verbal ability and student mathematics achievement. This is consistent with findings of studies of the relationship between teacher ability and student achievement without specific focus on mathematics and science.

## Teaching Experience

Studies generally measure teaching experience in terms of either teachers' total years of teaching or teachers' years of teaching in a given district. A few studies examine the impact of these measures on students' mathematics and science achievement. Ferguson (1991) found a positive significant relationship between years of experience and student achievement in reading and mathematics. In the primary grades, five to nine years and nine or more years of experience showed about equal effects on student test scores; in the secondary grades, teachers having nine or more years of experience showed a stronger effect. Hawkins, Stancavage, and Dossey (1998) found statistically significant associations between teacher experience and students' mathematics achievement. Fourth- and eighth-grade students who were taught mathematics by teachers with more than five years of

experience had higher mathematics scores than students who were taught by teachers with five or fewer years of experience. Fetler (1999) found teaching experience to be significantly positively related to high school students' mathematics scores. Goldhaber and Brewer (1997b) also found teaching experience to be positively related to high school students' mathematics scores; however, the result was not statistically significant. Rivkin, Hanushek, and Kain (2005) reported that students of beginning teachers perform significantly worse than those of experienced teachers on mathematics achievement tests. In science, Druva and Anderson's (1983) meta-analysis of 65 studies found student outcomes in science positively related to teachers' experience; however, the relationship was not particularly strong.

Other studies examining the impact of teacher experience on student achievement report mixed or no results. Rowan, Correnti, and Miller (2002) found that teacher experience was positively related to student mathematics achievement for a cohort of students in Grades 3 to 6 but not for a group of students in Grades 1 to 3. Ferguson and Ladd (1996) found no significant associations between teachers with five or more years of experience and third-, fourth-, eighth-, or ninth-grade students' mathematics achievement. Hill, Rowan, and Ball (2005) found no relationship between years of teaching and first and third graders' mathematics achievement.

In general, studies examining the relationship between teachers' years of experience and their effectiveness report somewhat mixed results. However more studies report a positive relationship (Ehrenberg & Brewer, 1995; Ferguson, 1991; Fetler, 1999; Goldhaber & Brewer, 1997b; Greenwald et al., 1996; Hanushek, 1996). Studies focusing on mathematics teachers and student achievement show similar results, although the results appear more consistent at the secondary level. This review found few studies reporting the relationship between teachers' experience and science achievement.

## Pedagogical Knowledge

Teacher education research often examines measures of teachers' pedagogical knowledge as an indicator of teacher quality. These studies use measures such as degrees in education, educational coursework, and scores on exams measuring professional knowledge. Studies of mathematics and science teachers' pedagogical knowledge have reported positive effects of education training on teachers' knowledge and practices (e.g., see Adams & Krockover, 1997; Gess-Newsome & Lederman, 1993; Valli & Agostinelli, 1993). Studies examining the relationship between degrees in education as a measure of teachers' pedagogical knowledge and student outcomes have been more mixed. Hawkins et al. (1998) found that students of fourth-grade teachers who had a college major in education or mathematics education significantly outperformed students of teachers with a major in



a field other than education, mathematics education, or mathematics. However, eighth-grade students of teachers who had majored in education did not perform as well as those who had majored in mathematics. Goldhaber and Brewer (2000) found that teachers with education degrees had no impact on high school students' science achievement and had a statistically significant negative impact on high school students' mathematics achievement.

Studies using coursework in education as a measure of teachers' pedagogical knowledge indicate a positive relationship between this training and student achievement, particularly at the secondary level. Druva and Anderson's (1983) meta-analysis indicated small positive correlations (coefficients less than .20) between K-12 student outcomes and teachers' background in science and education courses. Examining characteristics of a subgroup of emergency-certified secondary mathematics and science teachers in the NLES:88 data, Darling-Hammond, Berry, and Thoreson (2001) found that secondary students of emergency-certified teachers who had more education training had significantly higher achievement levels than students of teachers with less training.

### *Coursework in Subject-Specific Pedagogy*

The impact of courses taken in subject-specific pedagogy (i.e., mathematics education or science education methods courses) has also been examined. Chaney (1995) found that eighth-grade students whose teachers had taken coursework in both advanced mathematics (higher than calculus) and mathematics education had the highest mean standardized scores on NLES:88 mathematics test; students of teachers who had taken neither class of courses had the lowest mean standardized score. Chaney found no relationship between a background in science pedagogy and student achievement. Monk (1994) found that courses in undergraduate mathematics pedagogy contributed more to secondary students' achievement gains than did undergraduate mathematics coursework. In science, the study found coursework in science pedagogy positively related to secondary students' achievement, although these effects were much smaller. At the elementary level, Guarino, Hamilton, Lockwood, and Rathbun (2006) found no statistically significant relationship between kindergartners' achievement gains in mathematics and teachers' coursework in mathematics teaching methods.

Generally, much of the research on teachers' pedagogical knowledge examines the impact of teacher training on the development of teaching-related knowledge and skills. A few studies of mathematics and science teachers' pedagogical knowledge examine the impact of education degrees and coursework on student achievement. These studies indicate a more positive impact of degrees in education at the elementary level. At the secondary level, although there is evidence that education degrees have little or negative impact on student achievement, studies

indicate that coursework taken in subject-specific pedagogy is positively related to secondary student achievement, particularly in mathematics.

### Mathematics and Science Subject Matter Knowledge

Subject-matter knowledge is another teacher characteristic presumed to be indicative of teacher quality. Reviews of research indicate links between teachers' subject-matter preparation and student achievement, although these results are not always clear (Darling-Hammond, 2000; Darling-Hammond & Youngs, 2002; Wilson & Floden, 2003; Wilson et al., 2001). Most results that do show consistent positive links between subject matter knowledge and student achievement appear in the area of mathematics (Wilson & Floden, 2003). Common variables used to measure teacher subject knowledge include subject-specific degrees and coursework.

#### *Subject-Specific Degrees*

The results of studies examining the relationship between teachers holding subject-specific degrees and student achievement vary although mathematics results are generally positive, particularly at the secondary level (Chaney, 1995; Goldhaber & Brewer, 1997a, 2000; Rowan, Chiang, & Miller, 1997). For example, Goldhaber and Brewer (1997a, 1997b) found that teachers' holding bachelor's or master's degrees in mathematics had a statistically significant positive relationship to high school students' mathematics achievement (compared to teachers without advanced degrees or out-of-subject degrees). In science, they found holding a bachelor's degree in science (rather than having no degree or a BA in another subject) to have a statistically positive relationship with student achievement (Goldhaber & Brewer, 1997a). A later study found similar positive results for teachers' having a mathematics BA or MA on secondary students' mathematics achievement but no significant relationship between a science degree and secondary students' science achievement (Goldhaber & Brewer, 2000). Further, the studies found negative or little impact of teachers having nonsubject specific degrees on student achievement in mathematics and science. Using NLES:88 data, Rowan et al. (1997) found a positive association between teachers holding a degree in mathematics and Grade 10 students' mathematics achievement, although the effect was small. Chaney found significantly positive associations between teachers' having an undergraduate or graduate degree in mathematics and eighth-grade students' performance on the NELS:88 mathematics exam. In science, the same study found positive associations between eighth-grade students' science achievement and teachers' holding graduate degrees in science. Monk (1994), however, found no impact of a major in mathematics on secondary students' mathematics achievement but did find a significant positive relationship of a science major for junior year students' science achievement.

At the elementary level, studies generally focus on mathematics and indicate mixed or negative effects of teachers having subject-specific degrees on student achievement (Hawkins et al., 1998; Monk, 1994; Rowan et al., 2002). As previously discussed, Hawkins et al. found eighth-grade students whose teachers had majored in mathematics scored higher on the National Assessment of Educational Progress (NAEP) mathematics assessment. Yet researchers found no difference in mathematics performance between fourth-grade students whose teacher had majored in mathematics and students whose teacher had majored in education. Rowan et al. (2002) found that being taught by a teacher with an advanced degree in mathematics was negatively associated with mathematics achievement for elementary students. The researchers note that very few of the teachers in the sample had subject-matter degrees.

### *Subject-Specific Coursework*

Other studies measure teachers' subject matter knowledge using undergraduate or graduate coursework. Eisenberg (1977) found no significant relationships between algebra teachers' coursework in advanced mathematics, collegiate mathematics grade point average, scores on an algebra test, and student achievement. Chaney (1995), however, found that a background in advanced mathematics courses predicted eighth-grade student achievement in mathematics after controlling for teaching assignments. In science, teachers' subject area grade point average and having taken more than 40 credits in earth and physical sciences predicted student achievement. Druva and Anderson (1983) found student achievement to be positively related to the number of biology courses taken (for biology teachers) and the number of science course taken, in general. Monk (1994) found that the effects of teacher content preparation appeared to vary for different groups of students. For example, the number of mathematics courses in a teacher's background had a positive effect on students enrolled in advanced mathematics courses but no effect on students enrolled in remedial courses. Although the data suggested a positive relationship between coursework and students' mathematics achievement, there was evidence of a curvilinear effect in which the positive effect of a teacher's undergraduate subject coursework on student achievement diminished after five courses (Monk, 1994). In science, the effects of subject matter coursework were dependent upon the area of science studied (i.e., physical, earth, or life sciences). For example, Monk and King (1994) found that coursework in the life sciences had no impact on student achievement, but coursework in the physical sciences had a positive impact on higher ability students during the sophomore year. At the elementary level, Eberts and Stone (1984) found the number of college-level mathematics courses teachers had taken in the last three years was not significantly related to fourth-grade students' mathematics achievement gains.



### *Subject-Specific Knowledge for Teaching*

In the past two decades, researchers in teacher education have examined the nature of the knowledge needed for teaching and the role of this knowledge in teacher quality. This discussion stems from a perspective that knowing a subject for oneself is not adequate to effectively carry out the work of teaching. Rather, teachers must have an understanding of content as well as knowledge of how students think and understand the content. In other words, teachers use subject-specific content in their work differently from the way others might use content in nonteaching professions. Shulman (1986) introduced *pedagogical content knowledge*, knowledge “which goes beyond knowledge of subject matter per se to the dimension of subject matter knowledge *for teaching*” (p. 9). This knowledge includes ways of representing a topic in a way that makes is accessible to learners and understanding what facilitates or hinders learning of a topic. Other work in the area of subject knowledge for teaching has proposed various organizational structures for and theories of teacher knowledge (Ball, 1991; Ball & Bass, 2000; Grossman, 1990; Leinhardt & Smith, 1985; Ma, 1999). Much of this theoretical work has occurred in the area of mathematics. For example, Ma described *profound understanding of fundamental mathematics*, in her comparison of the nature of the subject matter knowledge of U.S. and Chinese elementary teachers. Such knowledge includes connections among topics and knowledge of multiple representations and explanations of topics. Ball (1991, 2003) described *mathematical knowledge for teaching*, which argues that teachers must not only know the subject matter for themselves but also understand the subject in a way that enables them to effectively use in it instruction.

Research in this area has focused on examinations and comparisons of preservice and expert teachers’ content knowledge for teaching (Ball, 1990; Leinhardt & Smith, 1985; Simmons et al., 1999; Stacey et al., 2001), changes in preservice and inservice teachers knowledge through participation in methods courses and professional development experiences (Borko et al., 1992; Davis & Krajcik, 2005; Kinach, 2002; Smith, 2000), international comparisons of teachers’ knowledge (An, Kulm, & Wu, 2004; Ma, 1999) and how teachers’ content knowledge for teaching might influence their instructional decisions and practices (McDuffie, 2004; Thompson & Thompson, 1996). Research on the relationship between teachers’ knowledge for teaching and student achievement has been limited and generally in the area of mathematics. Carpenter, Fennema, Perterson, and Carey (1988) examined relationships between teachers’ pedagogical content knowledge and elementary student achievement. The researchers used instruments designed to measure 40 first-grade teachers knowledge of children’s solutions of addition and subtraction problems including distinctions among problem types, knowledge of children’s strategies, and knowledge of their own students. They found teachers’ ability to predict whether their own students could solve different problems was



significantly correlated with their students' achievement in number facts and problem solving. Hill, Schilling, and Ball (2004) developed measures for the purpose of determining growth in teachers' mathematical knowledge for teaching. The researchers argue that up to this point, "scholars have not attempted to measure teachers' knowledge for teaching in a rigorous manner and thus cannot track its development or contribution to student achievement" (p. 14). Hill et al. (2005) used their measures to examine relationships between teachers' mathematical knowledge for teaching and first- and third-grade students' gains in mathematics. They found that teachers' knowledge was significantly related to student achievement gains in both grades.

Although the research is not definitive, studies indicate a trend toward a positive relationship between secondary teachers' subject knowledge and student achievement, particularly in mathematics. Secondary teachers who hold a bachelor's or master's degrees in mathematics appear to have positive impacts on student achievement. Results for science are similar, although not as strong. Teacher coursework in mathematics and science also appears to have a positive impact on student achievement, although at least one study found the impact diminishes after a particular number of courses and differs depending upon the level of course (remedial vs. advanced) in mathematics and the area of study (e.g., physical vs. life sciences) in science. At the elementary level, the impact of teachers' subject-specific degrees and coursework is unclear. However, the development of new measures of elementary teachers' mathematical knowledge for teaching holds promise of providing additional information on the relationship between teachers' knowledge and student achievement.

### Mathematics and Science Teacher Certification Status and Certification Routes

Teacher certification status is frequently used as a measure of the effects of knowledge gained from teacher preparation (Darling-Hammond, 2000; Darling-Hammond & Youngs, 2002). Comparisons are often made between those who are fully certified and those who hold provisional or emergency certification; several studies indicate an advantage in favor of fully certified teachers on measures of student achievement and teacher performance evaluations (Darling-Hammond, 2000; Fetler, 1999). Darling-Hammond (2000), using data from the 1993–94 School and Staffing Survey found a state's percentage of fully certified teachers (full certification and a major in their field) to be significantly and positively related to average NAEP mathematics scores in Grades 4 and 8. Mathematics achievement was negatively related to the percentage of all teachers less than fully certified, the percentage of all uncertified teachers new to the field, and the percentage of all uncertified new hires. Goldhaber and Brewer (2000) found that high school students of mathematics teachers with private school or no certification

did significantly worse on mathematics tests than students of teachers with in-field standard or emergency certification. Results for science were similar but not as strong. Guarino et al. (2006) found no statistically significant relationship between teachers' certification and kindergarteners' achievement gains in mathematics.

Several studies focusing on mathematics and science teaching and student achievement explore the impact of subject-specific certification. Hawk, Coble, and Swanson (1985) found that students in Grades 6 to 12 having teachers fully certified in mathematics scored significantly higher on achievement tests than students with out-of-field teachers, particularly in algebra. Goldhaber and Brewer (1997b) found a significant negative relationship between student achievement and certification (not subject specific) and a significant positive association between students' mathematics achievement and teachers' certification in mathematics.

Research also examines differences in the quality of regularly versus alternatively certified teachers, usually in terms of teacher outcomes such as subject area and professional knowledge tests, performance ratings, and teacher observations. Few studies were found examining associations between regularly and alternatively certified teachers and student outcomes. These are in the area of mathematics. Laczko-Kerr and Berliner (2002) compared Grades 3 to 8 students' SAT 9 reading, mathematics, and language arts scores (in 1998 and 1999) of matched pairs of certified and undercertified teachers, including teachers from Teach for America, a popular alternative certification program. Results indicated that students of certified teachers scored significantly higher in mathematics on the 1999 results. Results in 1998 were also positive but not significant. Further analysis within the group of undercertified teachers indicated that Teach for America teachers did not perform significantly differently than other undercertified teachers. Darling-Hammond, Holtzman, Gatlin, and Heilig (2005) found similar results.

Results of studies examining the relationship between teacher certification and student outcomes are often viewed as inconclusive because of wide variations in the scope and quality of such programs. However, studies using more targeted measures, such as subject-specific certification, report a positive relationship between certification status and student achievement, particularly in mathematics at the secondary level. The results for secondary science show similar trends but are generally weaker. There are fewer studies on the impact of subject-specific certification at the elementary level.

### Mathematics and Science Teacher Behaviors, Practices, and Beliefs

Teacher behaviors, instructional practices, and beliefs are also examined as indicators of teacher quality. Much of the research on the relationship between teachers' behaviors, practices, and beliefs and student outcomes occurs in mathematics. Peterson, Fennema, Carpenter, and Loef (1989) found a significant relationship

between first-grade teachers' pedagogical content beliefs about addition and subtraction and student achievement. Students of teachers who held a more cognitively based perspective (i.e., teachers who believe children construct their own knowledge, skills should be taught in relation to understanding, and mathematics instruction should build on students' prior knowledge and understanding) scored significantly higher on problem solving measures than students of teachers who held less cognitively based perspectives. A study examining teachers' pedagogical beliefs and elementary students' mathematics achievement found that students of teachers who held more constructivist beliefs did better on word problem tests than students whose teachers used a more direct-instruction approach (Staub & Stern, 2002). Carter and Norwood (1997) found a significant relationship between the alignment of fourth- and fifth-grade teachers' beliefs about mathematics to NCTM's *Principles and Standards for School Mathematics* and students' beliefs that working hard to solve challenging problems and understand concepts would lead to success. Stipek, Givven, Salmon, and MacGyvers (2001) found that fourth-through sixth-grade teachers' self-confidence as teachers of mathematics was significantly related to students' self-confidence as learners. Love and Kruger (2005) examined the relationship between teachers' beliefs and student achievement in urban elementary schools serving African American children. Results found a significant positive correlation between teacher beliefs that all children can be successful and students' achievement in mathematics.

Other studies examine relationships between mathematics teachers' practices and student outcomes. Analysis of data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998–99, found instructional practices of kindergarten teachers to be significantly associated with student gains in mathematics (Guarino et al., 2006). These include spending more time on the subject; using traditional instructional approaches; emphasizing computation; working on advanced numeracy, measurement, and other concepts; and the use of student-centered instruction techniques. Turner, Meyer, Midgley, and Patrick (2003) compared the relationship between differences in teachers' discourse and students' motivation in two sixth-grade classrooms. Students in the class with a higher occurrence of teacher discourse that encouraged student autonomy and fostered intrinsic motivation reported fewer instances of avoidance behavior or negative affect in the face of difficulties.

Research examining the relationship between teachers' use of instructional practices aligned with the NCTM *Standards* and reform-based curricula has found positive results. Sowell (1989) found positive relationships between the use of manipulative materials and K-16 student achievement. Stipek et al. (1998) found a positive relationship between teachers' instructional practices such as a focus on learning and understanding and fostering positive emotions toward mathematics learning and fourth- through sixth-grade students' achievement. Hiebert (1999), in a review of research, found a positive relationship between instructional



approaches that emphasize conceptual development of primary grades arithmetic and students' conceptual understanding of the topic. Ginsburg-Block and Fantuzzo (1998) found the use of problem solving and peer collaboration with low-achieving third and fourth graders positively associated with achievement in computation and word problem tasks. Cohen and Hill (2000) report a small positive relationship between California teachers' reported use of practices aligned with state mathematics reform and fourth-grade student achievement. Wenglinsky (2002, 2004) found instruction emphasizing higher order thinking, the use of hands-on learning, and solving problems with multiple solutions positively associated with fourth- and eighth-grade students' achievement on the mathematics NAEP. Hamilton et al. (2003) examined relationships between teachers' reported use of standards-based instruction (i.e., practices that support active learning, promote higher order thinking, and connect learning to real-world contexts) and student achievement in 11 K-8 schools. Although results indicated small, positive relationships to student achievement in mathematics, there were similar but not significant trends in science.

At the high school level, Mayer (1989) examined differences in achievement of middle and high school algebra students taught in classrooms using instructional practices aligned with the NCTM *Standards* and students in classrooms using more traditional approaches. Results indicated that students of teachers reporting greater use of standards-based practices had higher achievement growth than students of teachers reporting lower use of such practices. This was particularly true for higher ability students. An examination of high school classrooms using a standards-based curriculum found that teacher practices aligned with the goals of the curriculum (i.e., collaborative planning among teachers, collaborative work among students, use of multiple assessment methods, and emphasis on high expectations) were significantly related to growth in student achievement (Schoen, Cebulla, Finn, & Fi, 2003).

In science, research has examined the relationship between instructional practices that engage students in developing models, explaining and justifying claims, designing and conducting inquiries, and making use of meaningful problems and student outcomes (Committee on Science Learning, 2007). Kolodner et al. (2003) found the use of project-based inquiry approaches to have a positive impact on middle school students' learning. Students in the project-based class performed significantly better than students in the traditional class on collaborative, metacognitive, and science skills (e.g., designing tests and explaining and justifying claims). Rivet and Krajcik (2004) found that students in classrooms using project-based instruction showed large, significant gains in science content and process skills. Marx et al. (2004) found similar results for students in Grades 6, 7, and 8. Analysis of data from the National Educational Longitudinal Study found the use of hands-on laboratories positively related to 10th-grade students' science achievement (Burkam, Lee, & Smerdon, 1997). An examination of the



use of inquiry-based teaching practices found significant relationships between the use of these strategies and secondary students' science achievement (Von Secker, 2002). Teacher practices included encouraging students' interest in science, engaging students in laboratory and problem solving tasks, promoting students' further study of topics, and using scientific writing.

The research indicates that what mathematics and science teachers believe about teaching and learning and what they do in their classrooms have an impact on student outcomes. Much of the research uncovered for our review focused on the use of reform- or standards-based practices in mathematics instruction at the K-8 level. This research indicates positive results on student achievement. Studies focusing on mathematics at the secondary level indicate similar results. In science, studies indicate evidence of a positive relationship between the use of hands-on activities and practices related to inquiry-based instruction.

## DISCUSSION

The quality of mathematics and science teaching has been the focus of much attention in recent years as the United States faces the challenge of maintaining its competitiveness in an increasingly global economy (National Academies, 2006). International comparisons, such as the Trends in International Mathematics and Science studies, have indicated that U.S. students lag behind their peers in other countries in mathematics and science (Hiebert et al., 2003). As a result, the interest in identifying characteristics that determine quality in mathematics and science teachers has grown. The teacher quality research reviewed here, with a focus on mathematics and science, provides some insight into the relationships between teacher characteristics and student outcomes.

### What Do We Know About Mathematics and Science Teacher Quality?

This article indicates trends toward positive relationships between subject matter preparation (as measured by subject-specific degrees and coursework) and student achievement, particularly in secondary mathematics. Although the impact of non-specific degrees on secondary student achievement in mathematics and science has been inconclusive, evidence points to the generally positive impact of subject-specific degrees on secondary students' mathematics achievement (Chaney, 1995; Goldhaber & Brewer, 1997a, 1997b, 2000). In science, the research indicates that the relationship between science teachers' subject matter preparation and student achievement depends upon the area of science (e.g., physical science, life science, earth science, etc.). Although the relationship in science is less clear, there remains evidence of a positive trend. These results align with the emphasis on subject-matter preparation evident in the NCLB Act and related policy documents.

Although the evidence supports the idea that mathematics and science teachers must know their subject, there are indications that preparation in pedagogy is also beneficial. In mathematics, there is evidence of a positive relationship between subject-specific certification (which often includes both work in content and pedagogy) and student achievement at the secondary level (Goldhaber & Brewer, 1997b; Hawk et al., 1985). Evidence of positive associations (for secondary mathematics and, to a lesser degree, science) to coursework in subject-specific methods also supports this view. This conclusion closely aligns with recommendations by leading professional organizations such as the Council of Scientific Society Presidents, the NCTM, and NSTA, described previously, as well as other recent recommendations that advocate a closer link between mathematics content and pedagogy in the preparation of teachers (Conference Board of the Mathematical Sciences, 2001; Ferrini-Mundy & Findell, 2001).

Although the findings regarding mathematics and science subject matter preparation are generally positive at the secondary level, the impact of such preparation on the effectiveness of elementary teachers is inconclusive. Studies that examine the effect of subject-specific degrees and certification at the elementary level have noted the small number of teachers in the population who possess such credentials (Rowan et al., 2002); rather, elementary teachers are usually generalists and their credentials reflect this status. However, the focus on improving the quality of mathematics and science teaching and learning extends to K-12 and beyond. Therefore there is interest in determining what impact subject matter preparation might have on elementary teachers' effectiveness. Because subject-specific degrees and certification are not adequate measures for this teacher population, alternative measures are needed to determine how much and what type of subject-specific knowledge might be important. Ball (2003) argued that requiring teachers to study more mathematics is helpful only if teachers are learning the mathematics in ways that will help them help their students learn more mathematics. Research using instruments designed to measure mathematical knowledge used in teaching indicates that elementary teachers' performance on these measures are positively associated with student achievement in mathematics (Hill et al., 2005). Additional research using these and similar measures will further illuminate this issue.

### What Further Research on Mathematics and Science Teacher Quality Can Offer

Similar to reviews of research on general teacher quality (Wilson et al., 2001), the review of research on mathematics and science teacher quality and its relationship to student outcomes highlights the need for more targeted measures. For example, researchers need more information about the relationship between specific experiences and courses in mathematics and science teacher preparation and teacher quality in terms of student outcomes. What are the specific components of quality

mathematics and science teacher certification programs (whether traditional or alternative) that are positively related to student outcomes? What is the form and content of subject knowledge that most contributes to student learning? These questions would benefit from further exploration.

In addition, this review found fewer studies examining relationships between characteristics of teachers and student outcomes in the area of science than in mathematics. Those studies that examined relationships in both areas (Chaney, 1995; Goldhaber & Brewer, 1997a, 2000; Monk, 1994) often found more mixed results. One reason for this could be that the measures need to be refined to examine specific areas of science (e.g., physical, life, earth, and space sciences). For example, although examinations of subject-specific certification looked at secondary teachers' certification in science, they did not separate these into specific areas of certification and achievement such as biology, chemistry, or physics. Perhaps refining these measures to reflect topics in science education would yield more straightforward results. Because of the documented shortage of certified mathematics and science teachers (Ingersoll, 2001), these issues are becoming increasingly important.

More targeted measures will also give greater insights into the specific contributions of teacher characteristics that already appear to be important indicators of teacher quality. As seen in the example of degrees and certification, using measures of subject-specific degrees and certification provides more insight on the impact of these characteristics. Yet a mathematics degree from one institution is not the same as it is from another; similar variations exist in certification requirements across states. Further, as discussed earlier, the use of subject-specific degrees and certification may not be practical for some populations of teachers. Researchers need similar measures of teacher knowledge that can be used across several studies (Floden & Meniketti, 2005). This would provide more precise information on the impact of specific types of teacher knowledge on student outcomes and allow schools of education and other preparers of teachers to be more focused and targeted in the development and delivery of their programs. The shortage in qualified mathematics and science teachers must be addressed, and the development of alternative paths to certification is one response to this issue. As a result, there will continue to be an increase in the variety of available routes to becoming a teacher. Thus it is more important than ever to understand exactly what characteristics define a qualified mathematics and science teacher and how best to prepare an individual to fill that role.

## CONCLUSION

This article was developed to provide various audiences with an overview of the research on the relationships between individual characteristics of mathematics and science teachers and student outcomes. This is in no way an exhaustive

review of the entire body of literature on teacher quality. There exists much other research about relationships between the variables explored in this review and teacher outcomes (e.g., changes and influences on teacher beliefs, practices, and organization of knowledge), the effects of interventions designed to influence teacher quality (e.g., professional development experiences), and contextual issues in teaching and learning (e.g., school setting and student diversity). Although it was beyond the scope of this review to examine all of these areas, this broad synthesis of mathematics and science teacher quality as it relates to student outcomes is one resource for policymakers, educators, and researchers as they consider the complex issue of teacher quality in mathematics and science.

### ACKNOWLEDGMENT

This study is one in a series of substudies for the Math and Science Partnership Program Evaluation (MSP-PE) conducted for the National Science Foundation's MSP Program. The MSP-PE is conducted under Contract No. EHR-0456995. Since 2007, Bernice Anderson, Ed.D., Senior Advisor for Evaluation, Directorate for Education and Human Resources, has served as the National Science Foundation Program Officer. The authors, are Johnna J. Bolyard, from West Virginia University, and Patricia S. Moyer-Packenham, Utah State University (formerly of George Mason University).

The MSP-PE is led by COSMOS Corporation in current partnership with George Mason University (GMU) and Brown University. Robert K. Yin (COSMOS) serves as Principal Investigator and Jennifer Scherer (COSMOS) serves as one of three Co-Principal Investigators. Additional Co-Principal Investigators and their collaborating institutions (including discipline departments and math centers) are Patricia Moyer-Packenham (USU, formerly GMU) and Kenneth Wong (Brown).

### REFERENCES

- Adams, P. E., & Krockover, G. H. (1997). Beginning science teacher cognition and its origins in the preservice secondary science teacher program. *Journal of Research in Science Teaching*, 34, 633–653.
- An, S., Kulm, G., & Wu, Z. (2004). The pedagogical content knowledge of middle school mathematics teachers in China and the U.S. *Journal of Mathematics Teacher Education*, 7, 145–172.
- Association for Science Teacher Education. (2004). *ASTE Position Statement Science Teacher Preparation and Career-long Development*. Retrieved February 22, 2007, from <http://aste.chem.pitt.edu/>
- Ball, D. L. (1990). Prospective elementary and secondary teachers' understanding of division. *Journal of Research in Mathematics Education*, 21, 132–144.
- Ball, D. L. (1991). Research on teaching mathematics: Making subject matter knowledge part of the equation. In J. Brophy (Ed.), *Advances in research on teaching* (Vol. 2, pp. 1–41). Greenwich, CT: JAI Press.



- Ball, D. L. (2003). *What mathematical knowledge is needed for teaching mathematics?* Retrieved February 10, 2005, from the U.S. Department of Education Web site: <http://www.ed.gov/its/mathscience/ball.html>
- Ball, D. L., & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics. In J. Boaler (Ed.), *Multiple perspectives on the teaching and learning of mathematics* (pp. 83–104). Westport, CT: Ablex.
- Blank, R. K., & Langeson, D. (1999). *State indicators of science and mathematics education 1999: State-by-state trends and new indicators from the 1997-98 school year*. Washington, DC: Council of Chief State School Officers.
- Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational Researcher*, 34(6), 3–15.
- Borko, H., Eisenhart, M., Brown, C. A., Underhill, R. G., Jones, D., & Agard, P. C. (1992). Learning to teach hard mathematics: Do novice teachers and their instructors give up too easily? *Journal for Research in Mathematics Education*, 23, 194–222.
- Burkam, D. T., Lee, V. E., & Smerdon, B. A. (1997). Bender and science learning in early high school: Subject matter and laboratory experiences. *American Educational Research Journal*, 34, 297–331.
- Carpenter, T. P., Fennema, E., Peterson, P. L., & Carey, D. A. (1988). Teachers' pedagogical content knowledge of students' problem solving in elementary arithmetic. *Journal for Research in Mathematics Education*, 19, 385–401.
- Carter, G., & Norwood, K. S. (1997). The relationship between teacher and student beliefs about mathematics. *School Science and Mathematics*, 97, 62–67.
- Chaney, B. (1995). *Student outcomes and the professional preparation of eighth-grade teachers in science and mathematics*. Arlington, VA: National Science Foundation. (ERIC Document Reproduction Service No. ED 389530)
- Cochran-Smith, M., & Zeichner, K. M. (Eds.). (2005). *Studying teacher education: The report of the AERA panel on research and teacher education*. Mahwah, NJ: Erlbaum.
- Cohen, D. K., & Hill, H. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record*, 102, 294–344.
- Committee on Science Learning, Kindergarten through Eighth Grade. (2007). *Taking science to school: Learning and teaching science in grades K–8* (R. A. Duschl, H. A. Schweingruber, & A. W. Shouse, Eds.). Washington, DC: National Academies. Retrieved February 28, 2007, from [http://books.nap.edu/openbook.php?record\\_id=11625&page=R1](http://books.nap.edu/openbook.php?record_id=11625&page=R1)
- Conference Board of the Mathematical Sciences. (2001). *The Mathematical Education of Teachers—Issues on Mathematics Education* (Vol. 11). Providence, RI: American Mathematical Society. Retrieved February 22, 2007, from the Conference Board of the Mathematical Sciences Web site: [http://www.cbmsweb.org/MET\\_Document/index.html](http://www.cbmsweb.org/MET_Document/index.html)
- Council of Scientific Society Presidents. (2004). *The need for well-qualified science and math teachers*. Retrieved March 5, 2006, from the Council of Scientific Society Presidents Web site: [http://www.nctm.org/news/2004\\_12\\_cssp-position.pdf](http://www.nctm.org/news/2004_12_cssp-position.pdf)
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives*, 8(1). Retrieved February 2005 from <http://epaa.asu.edu/epaa/v8n1>
- Darling-Hammond, L., Berry, B., & Thoreson, A. (2001). Does teacher certification matter? Evaluating the evidence. *Educational Evaluation and Policy Analysis*, 23, 57–77.
- Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). *Does Teacher Preparation Matter? Evidence about Teacher Certification, Teach for America, and Teacher Effectiveness*. Retrieved August 16, 2005, from Stanford University, Center for Teacher Education and School Reform Web site: <http://schoolredesign.net/srn/news/certification.html>
- Darling-Hammond, L., & Youngs, P. (2002). Defining “highly qualified teachers”: What does “scientifically-based research” actually tell us? *Educational Researcher*, 31(9), 13–25.

- Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher*, 34(3), 3–14.
- Druva, C. A., & Anderson, R. D. (1983). Science teacher characteristics by teacher behavior and by student outcome: A meta-analysis of research. *Journal of Research in Science Teaching*, 20, 467–479.
- Eberts, R. W., & Stone, J. A. (1984). *Unions and public schools*. Lexington, MA: D. C. Heath.
- Ehrenberg, R. G., & Brewer, D. J. (1994). Do school and teacher characteristics matter? Evidence from high school and beyond. *Economics of Education Review*, 13(1), 1–17.
- Ehrenberg, R. G., & Brewer, D. J. (1995). Did teachers' verbal ability and race matter in the 1960s? Coleman revisited. *Economics of Education Review*, 14(1), 1–21.
- Eisenberg, T. A. (1977). Begle revisited: Teacher knowledge and student achievement in algebra. *Journal for Research in Mathematics Education*, 8, 216–222.
- Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation*, 28, 465–498.
- Ferguson, R. F. & Ladd, H. F. (1996). How and why money matters: An analysis of Alabama schools. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 265–298). Washington, DC: Brookings Institution.
- Ferrini-Mundy, J., & Findell, B. (2001). The mathematical education of prospective teachers of secondary school mathematics: Old assumptions, new challenges. In *CUPM discussion papers about mathematics and the mathematical sciences in 2010: What should students know?* (pp. 31–41). Washington, DC: Mathematical Association of America.
- Fetler, M. (1999). High school staff characteristics and mathematics test results. *Education Policy Analysis Archives*, 7(9). Retrieved from <http://epaa.asu.edu/epaa/v7n9>
- Floden, R., & Meniketti, M. (2005). Research on the effects of coursework in the arts and sciences and in the foundations of education. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education: The report of the AERA panel on research and teacher education*. (pp. 261–308) Mahwah, NJ: Erlbaum.
- Gess-Newsome, J., & Lederman, N. G. (1993). Preservice biology teachers' knowledge structures as a function of professional teacher education: A yearlong assessment. *Science Education*, 77, 25–45.
- Ginsberg-Block, M. B., & Fantuzzo, J. W. (1998). An evaluation of the relative effectiveness of NSTM standards-based interventions for low-achieving urban elementary students. *Journal of Educational Psychology*, 90, 560–569.
- Goldhaber, D. D., & Brewer, D. J. (1997a). Evaluating the effect of teacher degree level on educational performance. In W. J. Fowler (Ed.), *Developments in school finance, 1996* (pp. 197–210). Washington, DC: National Center for Education Statistics, U.S. Department of Education.
- Goldhaber, D. D., & Brewer, D. J. (1997b). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *The Journal of Human Resources*, 32, 505–523. Retrieved May 25, 2005, from JSTOR database.
- Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22(2), 129–146.
- Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, 66, 361–396.
- Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education*. New York: Teachers College Press.
- Guarino, C. M., Hamilton, L. S., Lockwood, J. R., & Rathbun, A. H. (2006). *Teacher qualifications, instructional practices, and reading and mathematics gains of kindergartners* (Report No. NCES 2006–031). Washington, DC: National Center for Education Statistics. (ERIC Document Reproduction Service No. ED491190)

- Hamilton, L. S., McCafrey, D. F., Stecher, B. M., Klein, S. P., Robyn, A., & Bugliari, D. (2003). Studying large-scale reforms of instructional practice: An example from mathematics and science. *Educational Evaluation and Policy Analysis*, 25, 1–29.
- Hanushek, E. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *The American Economic Review*, 61, 280–288.
- Hanushek, E. (1996). A more complete picture of school resource policies. *Review of Educational Research*, 66, 397–409.
- Hawk, P. P., Coble, C. R., & Swanson, M. (1985). Certification: It does matter. *Journal of Teacher Education*, 36(3), 13–15.
- Hawkins, E. F., Stancavage, F. B., & Dossey, J. A. (1998). *School policies and practices affecting instruction in mathematics: Findings from the National Assessment of Educational Progress*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement. (ERIC Document Reproduction Service No. ED 424116)
- Hiebert, J. (1999). Relationships between research and the NCTM standards. *Journal for Research in Mathematics Education*, 30, 3–19.
- Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., et al. (2003). *Teaching Mathematics in Seven Countries: Results from the TIMSS 1999 Video Study* (No. NCES 2003-013 Revised). Washington DC: U.S. Department of Education, National Center for Education Statistics.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42, 371–406.
- Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal*, 105, 11–30.
- Ingersoll, R. M. (2001). *Teacher turnover, teacher shortages, and the organization of schools: A CTP working paper*. Seattle: Center for the Study of Teaching and Policy. Retrieved February 28, 2007, from <http://depts.washington.edu/ctpmail/PDFs/Turnover-Ing-01-2001.pdf>
- Kinach, B. M. (2002). A cognitive strategy for developing pedagogical content knowledge in the secondary mathematics methods course: Toward a model of effective practice. *Teaching and Teacher Education*, 18, 51–71.
- Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., & Holbrook, J. (2003). Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting learning by design into practice. *The Journal of the Learning Sciences*, 12, 495–547.
- Laczko-Kerr, I., & Berliner, D. C. (2002). The effectiveness of "Teach for America" and other under-certified teachers on student academic achievement: A case of harmful public policy. *Educational Policy Analysis Archives*, 10(37). Retrieved March 4, 2005, from <http://epaa.asu.edu/epaa/v10n37/>
- Leinhardt, G., & Smith, D. A. (1985). Expertise in mathematics instruction: Subject matter knowledge. *Journal of Educational Psychology*, 77, 247–271.
- Leitzel, J. R. C. (Ed.). (1991). *A call for change: Recommendations for the mathematical preparation of teachers*. Washington, DC: Mathematical Association of America.
- Love, A., & Kruger, A. C. (2005). Teacher beliefs and student achievement in urban schools serving African American students. *Journal of Educational Research*, 99, 87–98.
- Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Erlbaum.
- Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Fishman, B., Soloway, E., Geier, R., et al. (2004). Inquiry-based science in the middle grades: Assessment of learning in urban systemic reform. *Journal of Research in Science Teaching*, 41, 1063–1080.
- Mayer, D. P. (1989). Do new teaching standards undermine performance on old tests? *Educational Evaluation and Policy Analysis*, 20, 53–73.
- McDuffie, A. R. (2004). Mathematics teaching as a deliberate practice: An investigation of elementary pre-service teachers' reflective thinking during student teaching. *Journal of Mathematics Teacher Education*, 7, 33–61.



- Mitchell, K. J., Robinson, D. Z., Plake, B. S., & Knowles, K. T. (2001). *Testing teacher candidates: The role of licensure tests in improving teacher quality*. Washington, DC: National Academy Press.
- Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, 13(2), 125–145.
- Monk, D. H., & King, J. A. (1994). Multilevel teacher resource effects in pupil performance in secondary mathematics and science: The case of teacher subject matter preparation. In R. G. Ehrenberg (Ed.), *Choices and consequences: Contemporary policy issues in education* (pp. 29–58). Ithaca, NY: ILR Press.
- National Academies. (2006). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Washington, DC: Author. Retrieved October 13, 2006 from the National Academies Web site: [http://books.nap.edu/openbook.php?record\\_id=11463&page=R3](http://books.nap.edu/openbook.php?record_id=11463&page=R3)
- National Board for Professional Teaching Standards. (2002). *What teachers should know and be able to do*. Retrieved October 16, 2006, from [http://www.nbpts.org/UserFiles/File/what\\_teachers.pdf](http://www.nbpts.org/UserFiles/File/what_teachers.pdf)
- National Commission on Mathematics and Science Teaching for the 21st Century. (2000). *Before it's too late: A report to the nation from the national commission on mathematics and science teaching for the 21st century*. Washington, DC: U.S. Department of Education. Retrieved February 20, 2007, from <http://www.ed.gov/inits/Math/glenn/report.pdf>
- National Commission on Teaching and America's Future. (1996). *What matters most: Teaching for America's future*. Retrieved July 8, 2006, from <http://www.nctaf.org/documents/WhatMattersMost.pdf>
- National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics. (2005). *Highly qualified teachers*. Retrieved April 25, 2006, from [http://www.nctm.org/about/position\\_statements/qualified.htm](http://www.nctm.org/about/position_statements/qualified.htm)
- National Research Council. (1996). *National science standards*. Retrieved February 22, 2007, from the National Academies Web site: <http://books.nap.edu/readingroom/books/nses/4.html>
- National Science Teachers Association. (2004). *Science teacher preparation*. Retrieved October 10, 2006, from <http://www.nsta.org/positionstatement&psid=42>
- Office of the U.S. Press Secretary. (2006, April 18). *Fact sheet: The American competitiveness initiative* (Press release). Washington, DC: Author. Retrieved February 21, 2007, from <http://www.whitehouse.gov/news/releases/2006/04/20060418-8.html>
- Peterson, P. L., Fennema, E., Carpenter, T. P., & Loef, M. (1989). Teachers' pedagogical content beliefs in mathematics. *Cognition and Instruction*, 6, 1–40.
- Rice, J. K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes*. Washington, DC: Economic Policy Institute.
- Rivet, A., & Krajcik, J. S. (2004). Achieving standards in urban systemic reform: An example of a sixth grade project-based science curriculum. *Journal of Research in Science Teaching*, 41, 669–692.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73, 417–458.
- Rowan, B., Chiang, F., & Miller, R. J. (1997). Using research on employees' performance to the study effects of teachers on students' achievement. *Sociology of Education*, 70(4), 256–284.
- Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale, survey research tells us about teacher effects on student achievement: Insights from the *Prospects* study of elementary schools. *Teachers College Record*, 104, 1525–1567.
- Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future academic achievement*. Knoxville: University of Tennessee Value-Added Research and Assessment Center.
- Schoen, H. L., Cebulla, K. J., Finn, K. F., & Fi, C. (2003). Teacher variables that relate to students achievement when using a standards-based curriculum. *Journal for Research in Mathematics Education*, 34, 228–243.



- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Simmons, P. E., Emory, A., Carter, T., Coker, T., Finnegan, B., Crockett, D., et al. (1999). Beginning teachers: Beliefs and classroom actions. *Journal of Research in Science Teaching*, 36, 930–954.
- Smith, D. C. (2000). Content and pedagogical content knowledge for elementary science teacher educators: Knowing our students. *Journal of Science Teacher Education*, 11, 27–46.
- Sowell, E. J. (1989). Effects of manipulative materials in mathematics instruction. *Journal for Research in Mathematics Education*, 20, 498–505.
- Stacey, K., Helme, S., Steinle, V., Baturo, A., Irwin, K., & Bana, J. (2001). Preservice teachers' knowledge of difficulties in decimal numeration. *Journal of Mathematics Teacher Education*, 4, 205–225.
- Staub, F. C., & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains: Quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology*, 94, 344–355.
- Stipek, D. J., Givven, K. B., Salmon, J. M., & MacGyvers, V. L. (2001). Teachers' beliefs and practices related to mathematics instruction. *Teaching and Teacher Education*, 17, 213–226.
- Stipek, D. J., Salmon, J. M., Givvin, K. B., Kazemi, E., Saxe, G., & MacGyvers, V. L. (1998). The value (and convergence) of practices suggested by motivation research and promoted by mathematics education reformers. *Journal for Research in Mathematics Education*, 29, 465–488.
- Strauss, R. P., & Sawyer, E. A. (1986). Some new evidence on teacher and student competencies. *Economics of Education Review*, 5, 41–48.
- Thompson, A. G., & Thompson, P. W. (1996). Talking about rates conceptually, part II: Mathematical knowledge for teaching. *Journal for Research in Mathematics Education*, 21, 2–24.
- Turner, J. C., Meyer, D. K., Midgley, C., & Patrick, H. (2003). Teacher discourse and sixth graders' reported affect and achievement behaviors in two high-mastery/high-performance mathematics classrooms. *Elementary School Journal*, 103, 357–382.
- U.S. Department of Education. (2002). *Meeting the highly qualified teachers challenge: The Secretary's annual report on teacher quality*. Washington, DC: U.S. Department of Education, Office of Postsecondary Education, Office of Policy, Planning, and Innovation. Retrieved October 4, 2006, from <http://www.ed.gov/about/reports/annual/teachprep/2002title-ii-report.pdf>
- Valli, L., & Agostinelli, A. (1993). Teaching before and after professional preparation: The story of a high school mathematics teacher. *Journal of Teacher Education*, 44, 107–114.
- Von Secker, C. (2002). Effects on inquiry-based teacher practice on science excellence and equity. *Journal of Educational Research*, 95, 151–160.
- Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, 73(1), 89–122.
- Wenglinsky, H. (2002). How schools matter: The link between teacher classroom practices and student achievement. *Educational Policy Analysis Archives*, 10(12). Retrieved February 24, 2007, from <http://epaa.asu.edu/epaa/v10n12/>
- Wenglinsky, H. (2004). Facts of critical thinking skills? What NAEP results say. *Educational Leadership*, 62(1), 32–35.
- Wilson, S., & Floden, R. E. (2003). *Creating effective teachers: Concise answers for hard questions. An addendum to the report "Teacher Preparation Research: Current Knowledge, Gaps, and Recommendations."* Denver, CO: Education Commission of the States. (ERIC Document Reproduction Service No. ED 476366)
- Wilson, S. M., Floden, R., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations. A research report prepared for the U.S. Department of Education*. Seattle: Center for the Study of Teaching and Policy, University of Washington. Retrieved March 2005 from <http://www.ctpweb.org/>

# Does MSP Participation Increase the Supply of Math Teachers? Developing and Testing an Analytic Model

John H. Tyler and Svetla Vitanova  
*Brown University*

An important feature of the Math and Science Partnership (MSP) Program of the National Science Foundation is to increase K-12 student achievement in math and science by increasing the quality, quantity, and diversity of the nation's K-12 math and science teachers. Because the underlying supply of math and science teachers is never directly observed, the central premise of this article is that an examination of the extent to which the MSP Program might impact the quantity and quality of math and science teachers requires careful thought and modeling.

With that starting point, this study first develops a model that supports a premise that shifts in underlying supply can be inferred from shifts in the percentage of certified math teachers employed when (a) salaries are constrained to be below market clearing salaries and (b) uncertified or "out-of-field" certified teachers can compete as substitutes for certified math teachers. The study then tests the plausibility of the model using data from Texas and in so doing provides preliminary estimates of the extent to which a school or school district's MSP participation affected the supply of certified math teachers available to that school or district. The results, although inconclusive on the question of the labor supply effects of MSP participation by a school or school district, do suggest the reasonableness of the model for future work when more appropriate data will be available.

The National Science Foundation's (NSF's) Math and Science Partnership (MSP) Program was established in 2002 to integrate the work of higher education with K-12 to strengthen and reform mathematics and science education. Among its five main features, the program proposes to improve K-12 teacher quality, quantity, and diversity. This study is concerned with the extent to which the MSP Program might impact the quantity and quality of math and science teachers available to

our nation's K-12 schools. Specifically, this study focuses on the possibility that the MSP Program might impact the quantity of certified math teachers available to our nation's K-12 schools and thereby improve the quality of math education.

What are the mechanisms through which increasing the number of certified math and science teachers could strengthen math and science education in the nation? First, if the nation's school districts were already employing all of the certified math and science teachers that were needed, then a call for more teachers suggests that the MSP Program is primarily about reducing class size, which might improve math and science achievement. On the other hand, if schools currently face a shortage of certified math and science teachers, and if the void is filled with uncertified and out-of-field teachers, then increasing the quantity of certified math and science teachers could impact achievement through changing the mix of certified/uncertified teachers.

In this study, we presume that it is primarily this latter mechanism, replacing uncertified with certified teachers, that motivates the MSP goal of increasing the quantity of math and science teachers. This presumption starts with the fact that math and science are widely regarded as "hard to staff" areas relative to need. For example, a 2002 study by the National Center for Education Statistics found that 63% of high school students taking a physical science class had teachers who did not have a certification or major in some area of physical science, and 36% of high school students in math courses had teachers who lacked a certification or a major in math (National Center for Education Statistics, 2002). Of course, whether increasing the percentage of certified teachers will lead to stronger math and science education rests on the extent to which certified math and science teachers are better at promoting achievement than are uncertified and out-of-field teachers. The literature on this topic is somewhat mixed.

A recent study of the Houston school district found that certified teachers were consistently more effective at increasing elementary school student achievement in reading and math than were uncertified teachers, including teachers from the Teach for America (TFA) program (Darling-Hammond, Holtzman, Gatlin, & Heiling, 2005). Meanwhile, a study by Mathematica Policy Research, Inc. found that TFA teachers were no less effective than traditionally certified teachers in terms of impacting the reading achievement of elementary school students and were slightly more effective than traditionally certified teachers in promoting student achievement in math (Glazerman, Mayer, & Decker, 2006). The difference in the two results could be a result of the different research designs. In the Mathematica study TFA and regularly certified teachers were randomly assigned to classrooms, whereas this was not the case in the Darling-Hammond et al. study. If regularly certified teachers are systematically assigned to classrooms containing students with a greater potential for achievement growth, then we would expect the more positive certified teacher results in the Darling-Hammond article.

The most recent study comparing the effectiveness of certified and uncertified teachers uses data from the New York City school district. Using rich data that allow the authors to control for average classroom and school characteristics, Rockoff, Kane, and Staiger (2006) found no difference in the effectiveness of certified, uncertified, or alternatively certified elementary and middle school teachers in increasing yearly math or reading gains.

A different strand of research asks whether students learn more math when taught by teachers who have more “mathematical knowledge.” One set of studies in this area indicated that teachers with mathematics coursework, majors, and mathematics degrees—factors presumed to be proxies for more mathematical knowledge—improve student achievement more than teachers lacking these credentials (Goldhaber & Brewer, 2000; Monk, 1994; Rowan, Chiang, & Miller, 1997). A second set of studies used direct, rather than proxy, measures of teachers’ mathematical knowledge and found a consistently positive relationship between the math knowledge of elementary, middle, and high school teachers and the math learning gains of students (Rowan, Chiang, & Miller, 1997; Hill, Rowan, & Ball, 2005; Hill, 2007). A finding from Hill (2007) that is particularly relevant for our study is that middle school teachers who have certification to teach mathematics have higher levels of math knowledge than certified teachers who do not possess this subject-area certification. Hill also found that teachers with experience teaching math at the high school level had more mathematical knowledge than middle school teachers who lacked this experience. In summary, this area of research suggests that the math knowledge possessed by teachers is positively correlated with the math gains of their students and that teachers possessing a certification to teach math, especially when coupled with high school teaching experience, have higher levels of math knowledge than uncertified or out-of-field certified teachers.

Thus, although the literature is less than conclusive on the importance of teacher certification in general, there is evidence that teachers holding certification to teach mathematics may produce greater student learning gains in math than uncertified or out-of-field teachers. In addition, when viewed in the context of our study it should be kept in mind that most of the studies looking at the effectiveness of teacher certification on student learning gains focus on elementary or middle school teachers and students. As a result, it is not clear what we learn about importance of certification for the effectiveness of *high school* math teachers, the group under study here. Regardless of the research at this point, however, No Child Left Behind (2002) cites full state certification as one of the criteria for a teacher to be considered “highly qualified” in that landmark legislation. Given this, the federal government would apparently view an increase in the percentage of high school math teachers who are fully certified to teach mathematics as a potential way of improving student achievement.



Relative to that proposition, the purpose of this article is twofold. The first goal is to develop a tractable model of teacher supply and demand that can be informative regarding how shifts in the supply curve of certified teachers translate into changes in the employment levels of certified teachers when (a) salaries are not market determined and (b) uncertified or out-of-field teachers can compete for vacancies. The role of the model here is to provide a sound basis for making inferences about MSP Program impact on teacher quantity using observed employment levels of certified teachers.

The second goal of the article is to examine the model's potential usefulness as a tool for evaluating the MSP Program's role in increasing the quantity of math and science teachers available to our nation's schools. This is done in a preliminary manner here by applying the model to data on high school math teachers from Texas. We focus on Texas because at this point it is the first state where we have been granted ready access to data that contain the necessary information on teachers, schools, and school districts for testing the model.<sup>1</sup> We focus on math teachers in this article because, of the three MSPs that can be studied in Texas, two are targeted solely toward improvements in K-12 math, whereas the third is devoted to both math and science.<sup>2</sup>

We emphasize that the *primary* contribution of the article is the development and initial application of the model, not in the actual estimation of the impact of the MSP Program on the labor supply of math teachers in Texas. As is discussed later, given the data available for this research, it is likely too soon for the MSP Program to have a substantial impact on the labor supply decisions of teachers in Texas. However, the model illustrates that given appropriate data and sufficient time for the MSP Program to affect the decisions of teachers and potential teachers, one should be able to measure how the MSP Program impacts the labor supply curve of math and science teachers via changes in the number of certified math or science teachers employed by participating schools or school districts. The estimates that we do present later in the article are discussed only from a standpoint of whether

---

<sup>1</sup>The minimum data requirements are data that contain employment information by school (campus), by year, and by subject area on the number of (a) employed full-time equivalent certified teachers teaching in their field, (b) certified teachers who are teaching out of field, and (c) uncertified or alternatively certified teachers. The necessary data would have both MSP and non-MSP schools, preferably in the same state and would cover both pre- and post-MSP participation years. Although the Schools and Staffing Survey from the National Center for Education Statistics has the necessary teacher employment variables, the fact that this survey is not longitudinal means that we cannot make pre- and post-MSP comparisons, rendering this data set unsuitable for the purposes of this article. We are still searching for other potential sources of data.

<sup>2</sup>The Alliance for Improvement of Math Skills and the Texas Middle and Secondary Mathematics Project are targeted at math, whereas the El Paso Math and Science Partnership is concerned with both math and science. A fourth project, the Rice University Mathematics Leadership Project was developed at a latter date than the other three MSP project in Texas and so we do not consider that project in our analysis.

they support the model as a reasonable tool for a future examination of the impact of the MSP Program on teacher quantity.

## THE LITERATURE ON TEACHER SUPPLY

An examination of the literature on teacher supply points to the importance of salary in determining the quantity and quality of teachers but offers little in the way of modeling the labor supply of certified math and science teachers when salaries are determined by institutions, not markets, and when uncertified teachers can serve as substitutes for uncertified teachers. The most extensive work on teaching and the labor supply of teachers is found in *Who Will Teach? Policies That Matter* (Murnane et al., 1991). In this classic study of teacher supply and demand, attention is given to the shortage of math and science teachers in the nation's schools. The primary finding here emphasizes the importance of the salaries offered to math and science teachers relative to the out-of-teaching opportunities afforded to potential teachers with training in math and/or science. This work provides support for a key assumption in our supply/demand model, namely, that because of the way salaries are determined in public school districts, science and math teachers are offered below-market wages. The importance of salary, both starting salary and the potential for earnings growth over a lifetime in determining who enters and who stays in teaching, is echoed in other studies including Zabalza (1979), Manki (1987), and Dolton (1990).

Studying teachers in Arkansas, Galchus (1994) found that, along with salary, the characteristics of students played an important role in attracting well qualified teachers to a county's schools. Lankford, Loeb, and Wyckoff (2002) find that proximity to the area where one went to high school themselves is an important factor in understanding the sorting, and hence the supply, of teachers across districts in New York state. Stinebrickner (2001) used longitudinal data and a dynamic, discrete choice model to study the labor supply of certified elementary and high school teachers, analyzing the relationships between personal characteristics, wages, and the decision process of certified teachers. He found that important considerations in explaining exits from teaching in the first years after certification are marital and fertility decisions. The Stinebrickner study offers some support for the common notion regarding a "reserve pool" of certified teachers who are currently not teaching but could be induced to return to teaching. The actual size of the reserve pool of math and science teachers is undetermined, but various studies suggest that about one in four new hires each (all grades, all subjects) are from the reserve pool of certified teachers.<sup>3</sup> Other papers that study the factors that influence the decision to stay in or exit

---

<sup>3</sup>See, for example, Broughman and Rollefson (2000).

from teaching include Brewer (1996), Dolton and van der Klaauw (1995, 1999), Gritz and Theobald (1996), Mont and Rees (1996), , Murnane and Olsen (1989, 1990), Stinebrickner (1998), Theobald and Gritz (1996), and van der Klaauw (1996).

Some of the most recent and extensive work on the supply of and demand for mathematics and science teachers has been conducted by Richard Ingersoll of the Graduate School of Education at the University of Pennsylvania. Ingersoll's various analyses primarily draw on data from the National Center for Education Statistics' School and Staffing Survey and its supplement, the Teacher Follow-Up Survey. This work emphasizes the fact that teacher "shortages" can arise both because (a) teachers are not selecting into teaching jobs at rates required to meet demand and because (b) current teachers leave either the profession or their current teaching job for other opportunities (Ingersoll, 2000, 2003). His prescriptions for addressing the chronic shortage of math and science teachers focus more on the latter problem, with policies and practices designed to keep more current teachers in their jobs, rather than policies that focus on increasing the supply of math and science teachers. Factors that he cites include better teacher induction programs, improving student discipline, and giving teachers more influence over school policies that affect their jobs. Ingersoll argued that reducing exits from the profession and job transition within teaching address the shortage dilemma by reducing demand, rather than by increasing supply. To be noted, however, is that if the demand for math and science teachers in a given school or school district is relatively inelastic, then policies that reduce teacher exits from current positions can be viewed as shifting the supply curve rather than reducing demand. We develop this idea further when discussing our model of teacher labor supply and demand below.

In summary, a consistent and not surprising theme in the existing literature is that both starting salary and potential salary growth are important factors in the decisions of individuals to enter into and stay in teaching. Research also suggests that other factors such as proximity to home, marital and fertility decisions, and working conditions associated with teaching are also important. We draw on this literature in developing a supply–demand model for certified math teachers, and to help inform the specification of the resulting estimating equation based on that model.

In what follows, we argue that in the short run, teacher salaries are fixed by the district-level teacher salary schedule and therefore do not drive *within-school* differences over time in the supply of certified math teachers. Nevertheless, we control as best we can for the effects of teacher salary on the number of certified teachers employed. Also, because all of the estimates we present are from models that control for time-invariant factors at the campus level, many factors identified in the literature as important to explaining teacher labor supply decisions are accounted for. We need to consider potential differences between MSP and non-MSP schools for which we cannot control. In particular any unobserved,

*time-varying*, and systematic differences between the MSP and non-MSP schools or school districts that might be related to the employment of certified teachers could bias our estimates.

The organization of this article is as follows. In the next section, we model the supply of and demand for certified math teachers when salaries are not market determined and uncertified teachers can compete for jobs. Although the model is developed discussing math teachers, it is equally applicable to certified science teachers. In the Empirical Model for Estimating Changes in Labor Supply section, we present the estimating equation that is used to study the impact of MSP participation on teacher supply. In the Data section, we discuss the data we use, and in the following section we present the results. In the final section we conclude with a discussion of what is learned from the article.

## MODELING THE SUPPLY OF AND DEMAND FOR CERTIFIED TEACHERS

From an economic perspective, framing inquiry into the supply of certified math teachers begins with thinking about the supply and demand curves in the labor market for middle school and high school math teachers because elementary school teachers are not certified by subject area. Of course, both supply and demand curves are “if-then” propositions—if the price is  $Y$  then the supply will be  $X$ . In this setting, increasing the supply of math/science teachers devolves to two possibilities: price (salary) increases that cause movement along a static supply curve and/or “shocks” or changes in the labor market environment that cause an outward shift in the supply curve. The first involves drawing more teachers into the profession by, all else equal, raising salaries, whereas the second involves having a greater supply of teachers at any given salary. Because the MSP projects have little ability to affect teacher salaries, we model the impact of the MSP Program on the supply of math teachers via shifts the teacher supply curve rather than movements along the supply curve.<sup>4</sup>

In considering individual decisions regarding whether to teach, we focus on the fact that individuals considering any profession and the jobs therein are considering a “bundle” of attributes that are associated with professions and jobs. The job “bundle” consists of at least the following elements: starting salary and anticipated salary growth, the direct and opportunity costs of training and (if required) certification for a given profession, the probability of finding employment in that profession upon completion of the training, expected working conditions, distance from home to work, demands of family, safety, the intrinsic satisfaction

---

<sup>4</sup>We note that there is one MSP that entered into negotiations with the local teachers’ union to change teachers’ salaries.



associated with the work, and career wage and job quality trajectory. Individuals who are considering teaching think about the teaching bundle compared to the bundle associated with their best alternative, an alternative that may be either a job out of teaching or a decision to stay out of the labor force altogether. Thus, individuals on the margin in their decision of whether to teach relative to the best alternative can be induced to teach by manipulating any of the attributes associated with teaching as a profession and/or manipulating the attributes associated with a particular teaching job. The important ideas here are that (a) individuals are always comparing teaching to the best alternative use of their time, and (b) different individuals will place different weights on the differing attributes associated with the profession and any particular job.

### MSP Participation and the Labor Supply Decisions of Teachers

Given the concept of a weighted bundle of profession and job attributes, what are the mechanisms through which a school or school district's participation in the MSP Program could affect the available supply of certified teachers for that school or district? First, participating in the MSP Program could signal a certain kind of attractive professional environment to newly certified teachers and individuals from the reserve pool of certified, nonteaching individuals. For example, new and reserve-pool certified teachers who value a school environment where professional knowledge and intellectual growth are supported and encouraged might be attracted to MSP schools if they had information that a school was focusing on these elements via their MSP-related activities. Also, if certified teachers felt that MSP schools supported policies and practices that would make the job of teaching easier and more rewarding, they might be attracted to MSP schools. In general, anything that MSP participation does that translates into an MSP school being perceived as a better place to teach math would, in theory, attract more certified math teachers, holding salaries at the school constant.

There is a second mechanism through which MSP participation could affect the supply of certified teachers available to a school. By design the MSP Program establishes partnerships between K-12 schools and institutions of higher education (IHEs), most of which have teacher preparation and certification programs. These partnerships with IHEs could lead to an increased supply of certified math teachers available to the MSP K-12 schools by strengthening the linkages between the MSP school and newly minted certified teachers from the partnering IHE or between the MSP school and reserve pool of certified teachers who might have ties to the IHE.

Thus, there are at least two general mechanisms through which MSP participation could shift the labor supply of available certified math teachers: (a) MSP-related policies, practices, and activities that make MSP schools more attractive places to work and (b) the establishment or tightening of linkages between

MSP schools and partnering IHEs that produce certified math teachers. The question is, To what extent would we expect an increased supply of certified teachers available to a school to manifest itself in the increased employment of certified math teachers in a situation where (a) potential teachers value salaries, (b) salaries are constrained to be below market clearing levels, and (c) uncertified teachers are readily available and can substitute for certified teachers?

### A Simple Labor Supply Model Under Constrained Salaries

To answer this question, we consider the supply and demand for certified high school math teachers in a given school district as captured in Figure 1.<sup>5</sup> Figure 1 summarizes the markets for certified and uncertified high school math teachers facing a given school district.<sup>6</sup> Points on the rightward-pointing  $U$ -axis represent the number of uncertified teachers hired in the district, whereas the vertical  $C$ -axis represents the number of certified teachers hired.<sup>7</sup> The  $W_C$  axis to the left reflects the wage (salary) offered to certified teachers and the  $W_U$  axis depicts the wage offer to uncertified teachers. Given this setup, the line labeled  $D$  in the upper left quadrant is the district's demand curve for certified teachers as a function of the wage offer and  $S$  is the supply of certified math teachers as a function of wage. The line labeled  $S_U$  in the lower right quadrant is the supply curve of uncertified teachers as a function of wage.<sup>8</sup> The horizontal  $S_U$  supply curve for uncertified teachers reflects the assumption that within some relevant range, there is an unlimited number of uncertified teachers willing to work at a wage of  $w_u^k$ .

The line  $D_U(c)$  in the upper right quadrant is the district's demand curve for uncertified math teachers as a function of  $c$ , the number of certified teachers it can hire. If the district could hire amount  $A$  certified teachers, it would hire no uncertified teachers. If it could hire no certified teachers, it would have to hire all uncertified teachers out to  $B$ . In the short run we assume no substantial increase or decrease in the number of total students, and hence total math teachers employed, as a function of quality. We also assume that in the short run the district cannot or

<sup>5</sup>In both the modeling and the empirical analysis that follow, the focus is on high schools, because it is at the high school level that teachers are hired and appear in administrative records by subject.

<sup>6</sup>For narrative simplicity we use the term *uncertified* to represent both of the following categories: math teachers who lack standard certification and certified teachers who are teaching math as an out-of-field subject.

<sup>7</sup>Quantities along all dimensions in Figure 1 are increasing in the direction of the arrows of the axes.

<sup>8</sup>We note that the elasticity of supply of certified teachers with respect to salary is important, as the more elastic the supply, the greater the ability of a shift in supply to show up in employment levels. Given that we think that individuals with training in math and science are often "lured" away from teaching by better outside wage offers, the anecdotal evidence at least suggests that the supply of math and science teachers is relatively price elastic.

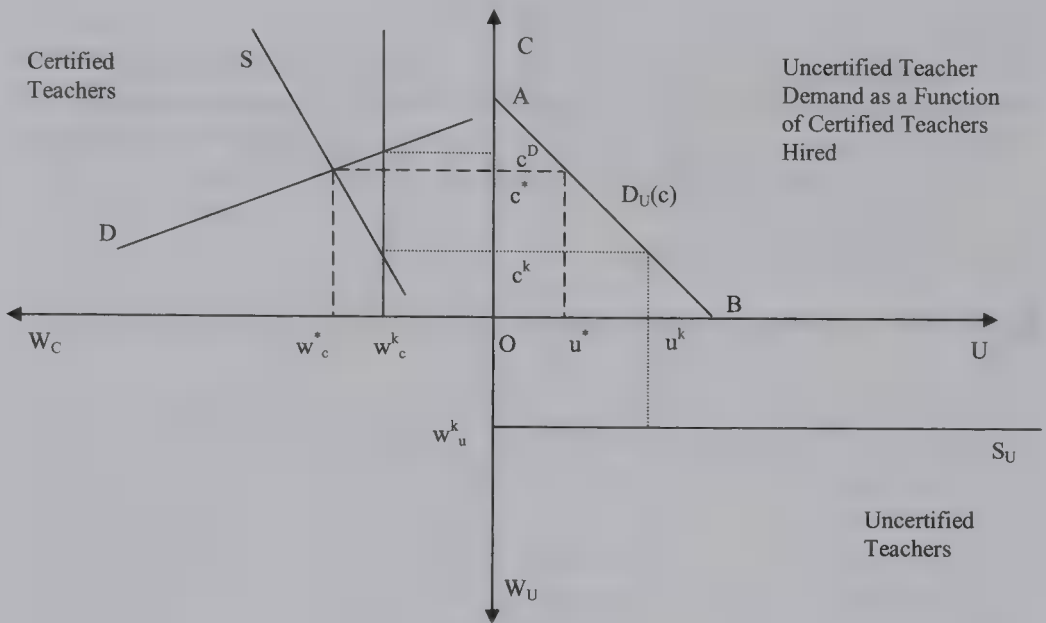


FIGURE 1 Supply and demand for certified and uncertified teachers in the short run.

does not adjust class size to offset any shortage of certified teachers. Rather, they hire uncertified teachers to fill their open math teaching positions, that is,  $OA = OB$ , and the slope of  $D_U(c)$  is  $-1$ .

To begin the analysis consider the equilibrium case where the school district paid the market clearing wage of  $w_c^*$  to certified math teachers. At this wage the supply of certified math teachers ( $c^*$ ) would equal district demand for certified teachers. Given demand for uncertified math teachers as a function of certified teachers who can be hired, the district will hire  $u^*$  uncertified teachers to fill out its math faculty. At the equilibrium wage the overall quality of the math faculty will be  $c^*/(c^* + u^*)$ , i.e., the proportion of math faculty in the district who are certified to teach math.

In fact, however, the salaries of all teachers in the district are determined by the district salary schedule that is primarily a function of experience and education beyond a BA degree. Because the salary schedule is *not* a function of subject matter taught, certified math teachers receive the same salary offer as all other teachers, conditional upon experience and education level.<sup>9</sup> Given the outside opportunities typically open to individuals with training in math, we assume that this constrained salary is below the market-clearing, equilibrium salary, and certified math teachers

<sup>9</sup>Although there are some few districts nationwide that offer an extra stipend to math or science teachers, these stipends tend to be low relative to base salary.





In summary, the short run results from a supply curve shift are that more certified math teachers are hired in the district *at the same teacher salary scale wage*, fewer uncertified teachers are hired, and teacher quality goes up since  $c_2^k + u_2^k = c^k + u^k$  and  $c_2^k > c^k$  so that  $c_2^k / (c_2^k + u_2^k) > c^k / (c^k + u^k)$ .

The key insight from Figures 1 and 2 is that although we cannot directly observe changes in the supply curve of certified teachers, we can make inferences about shifts in the supply curve through observed changes in the number of certified math teachers employed in the school district.<sup>11</sup>

To this point the analytics and the discussion have focused on the school district as the level of analysis. This is because it is usually at the district level that decisions regarding both salary levels and the number of teachers employed are made. In practice, however, it can be the case that in some districts only a portion of the schools are involved in an MSP partnership, whereas in other districts all of the schools are involved in the partnership. Given this reality, our empirical analysis presents results that use all of the Texas high schools that participated in an MSP partnership as well as results that use only those schools that are in school districts where *all* of the high schools are participating schools in the MSP project. This distinction is potentially important for interpreting any empirically based estimates for the following reason. For an MSP high school located in a district where not all high schools are participating in the project, there could be a positive MSP effect on the labor supply available to that MSP *school* that does *not* get translated into employment changes at that school because the school does not get to make the hiring decisions. Therefore, we would expect any estimated MSP effects to be smaller when estimated over a sample using all schools than when estimated using a sample where the MSP schools that are used are in districts fully “saturated” with MSP schools.

Another clarification is important at this point. In our study an MSP “participating” school is a school that has substantive involvement with the MSP Program. Our definition of “substantive involvement” derives from three variables in the MSP Program’s Management Information System (MSP-MIS). These data identified whether schools had met one of three conditions during school years 2002–2003 or 2003–2004:

- MSP-MIS item q5Bald: Whether 30% or more of targeted teachers participated in 30 or more hours of MSP-sponsored activities.

---

<sup>11</sup>If demand for education in the district is a function of teacher quality, then it is a simple matter to show that in the long run total teacher demand, which we are holding fixed, is a function of the ratio of certified to uncertified teachers and is therefore endogenous. Given the recent implementation of the MSP Program relative to available data, we focus on the short run analysis in this article.

- MSP-MIS item q5Bbld: Whether 30% or more of targeted students engaged in a challenging mathematics or science curriculum that was initiated or revised with MSP support.
- MSP-MIS item q5Bdld: Whether 30% or more of targeted students participated in an MSP-supported academic enrichment activity.

If at least one of the three conditions was met for either of the two years, we considered the school to be an MSP “participating” school.

An important caveat to our model and our approach of looking at certified teacher employment to make inferences about MSP effects is that district hiring practices tied to practices and policies such as seniority preference or union contracts may impede the substitution of certified for uncertified teachers in the event of an increase in the supply of certified teachers. We have no way of addressing this issue with our data and, as a result, our estimates based on number of full-time equivalent (FTE) teachers employed may be lower bound estimates of any actual shifts in the labor supply of math teachers. We do note that Texas is a right-to-work state, and so any downward bias would be less likely in Texas than in more unionized states.

Before leaving this section it is important to note what we are *not* directly examining. Namely, we are not studying whether IHEs that are in the MSP Program tend to produce more certified teachers than they otherwise would have. Although this is perhaps the most direct way that the MSP Program could increase the quantity of math teachers in the nation, we leave that question for another study because of lack of appropriate data at this time. We do note, however, that to the extent that the Texas MSP IHEs in this study produce more certified math teachers *and* to the extent that these newly minted teachers are attracted to the MSP high schools in the study, we will pick up that effect in our estimates.<sup>12</sup>

## EMPIRICAL MODEL FOR ESTIMATING CHANGES IN LABOR SUPPLY

In this section we use the model relating the supply of and demand for certified teachers as the basis for an empirical examination of the impact of MSP participation by a school or school district on the supply of certified math teachers in the local teacher labor market. Because the group of high schools that participate in the MSP Program is a nonrandom subsample, we use econometric methods in an

---

<sup>12</sup>Of course, a thorough study of the amount of newly certified high school math teachers that resulted from an IHE’s participation in the MSP Program would capture *all* newly certified teachers, both those who took employment in participating MSP high schools and those who took employment in non-MSP high schools.

attempt to control for confounding factors that could bias estimates of the causal impact of MSP participation on the employment of certified math teachers.

We have selected Texas for this study because of the availability of data on the number of FTE teachers by grade level, subject area, and certification status who were hired at every public high school in Texas for the academic years 2001–2002, 2002–2003, and 2003–2004.<sup>13</sup> For simplicity we refer to these years as 2002, 2003, and 2004, respectively. During these years there were 37 high schools across 19 school districts in Texas that were in one of the three MSP projects in that state, for which we were able to construct complete data and which met our criteria for being an MSP “participating” school. We also had complete data for 1,143 non-MSP Texas high schools over the 2002 to 2004 period. In this context 2002 is the “pre-program” year, and our analysis is primarily concerned with estimating whether MSP Program schools in Texas tended to employ more in-field, certified math teachers between the pre- and postprogram years than did observationally similar non-MSP participating schools in Texas. Drawing upon the lessons of the earlier Figures 1 and 2, we assume that conditional upon the prevailing salaries, most schools in Texas would like to hire more math teachers than are available given current supply, and thus any outward shift in the supply curve would be evidenced by that school hiring more certified math teachers, holding salary constant.

To fix ideas consider time periods, 0 and 1, representing before and after potential participation in the MSP Program, and the common coefficient model represented by the following two equations:

$$C_{ij0} = \alpha W_i + \lambda D_j + \gamma TM_{ij0} + \beta X_{i0} + \pi Q_{j0} + \delta MSP_{ij0} + u_{ij0}$$

$$C_{ij1} = \alpha W_i + \lambda D_j + \gamma TM_{ij1} + \beta X_{i1} + \pi Q_{j1} + \delta MSP_{ij1} + u_{ij1}$$

where  $i$  indexes schools,  $j$  indexes school districts, 0 and 1 index the pre- and post MSP periods, respectively, and  $C$  = number of FTE certified math teachers employed in school  $i$  and district  $j$ ,  $W$  = a vector of school level factors that do not vary between the two periods,  $D$  = a vector of district level factors that do not vary between the two periods,  $TM$  = total number of math teachers (certified, certified but teaching out of field, and uncertified) in school  $i$  and district  $j$ ,  $X$  = a vector of school level factors that may change between periods and may be correlated with both MSP participation and the number of certified teachers hired,  $Q$  = a vector of district level factors that may change between periods and may be correlated with both MSP participation and the number of certified teachers hired,  $MSP = 1$

<sup>13</sup>The different states of certification status in the data are certified and teaching in field, certified but teaching out-of-field, and uncertified.

if school  $i$  is an MSP participating school in the period of observation and zero otherwise, and  $u_0$  and  $u_1$  are mean zero, homoskedastic error terms.

Because  $MSP = 0$  for all schools in period 0, subtracting the two equations yields,

$$\Delta C_{ij} = \gamma \Delta TM_{ij} + \beta \Delta X_i + \pi \Delta Q_j + \delta MSP_{ij} + \varepsilon_{ij} \quad \text{with} \quad \varepsilon_{ij} = u_{ij1} - u_{ij0}$$

The parameter of interest in this differenced equation is  $\delta$ , a parameter that we would like to interpret as the causal impact of MSP participation on the change in the number of certified math teachers employed between periods 0 and 1. The differenced equation illustrates the fact that this inference is correct to the extent that we successfully control for those factors that change over time and that are correlated with both MSP participation and with the change in number of certified teacher hired.

We make special note of the inclusion of  $TM$ , the total number of employed math teachers in the previous equations. Controlling for the total number of certified and uncertified math teachers in the analysis accounts for the fact that participation in the MSP Program could affect demand as well as supply.

Given the data at hand in Texas, we estimate a school fixed-effects variant of the differenced equation just listed. The most correct specification will capture the effect of MSP participation on *percentage* changes in the number of math teachers rather than changes in absolute levels of teachers. Therefore, our analyses use the natural log of the number of certified math teachers rather than the number of teachers in levels. With these considerations, our primary estimating equation is

$$Y_{ijt} = Y03_t \beta_1 + Y04_t \beta_2 + Y03_t * MSP_{ij} \delta_1 + Y04_t * MSP_{ij} \delta_2 + TM_{ijt} \gamma + X_{it} \beta_3 + Q_{jt} \pi + \alpha_i + \varepsilon_{ijt} \quad (1)$$

where  $i$  indexes schools,  $j$  indexes school districts,  $t$  indexes time (year), and  $Y$  = natural log of the number of certified FTE math teachers employed by school  $i$  in district  $j$  in year  $t$ ,  $Y03_t$  = one if  $t = 2003$  and zero otherwise,  $Y04_t$  = one if  $t = 2004$  and zero otherwise,  $MSP$  = one if school  $i$  is an MSP school and zero otherwise,  $TM$  = natural log of the total number of math teachers employed in school  $i$  in year  $t$ ,  $X$  = a set of time-varying, school level covariates,  $Q$  = a set of time-varying, district level covariates,  $\alpha_i$  = time-invariant fixed effect for school  $i$ , and  $\varepsilon$  = the error term.

Equation 1 is fit on a stacked panel of data for the years 2002, 2003, and 2004. Robust standard errors are estimated and adjusted for the fact that schools are clustered within districts. In this specification, the variables  $Y03$  and  $Y04$  capture the main effects of the math teacher labor market in the program year periods relative to 2002, the period before participating schools entered into the MSP



Program. In this fixed effects model, all estimates represent *within-school* changes over time. Estimates of  $\delta_1$  capture the *within-school* percentage change in FTE certified math teachers employed by MSP Program schools between 2002 and 2003 relative to within-school percentage changes in certified math teachers employed over this period by non-MSP Program schools. Inferences about the impact of MSP participation on the employment of certified math teachers are unbiased to the extent that the non-MSP comparison schools in Texas accurately estimate what would have been the within-school employment practices of the MSP schools over the 2002 to 2003, conditional upon the other variables in the equation.

Because 2003 was the first year of MSP participation, one might expect any impact on the hiring practices of participating schools to be minimal. In this case interest is focused on estimates of  $\delta_2$ . This parameter captures the percentage change in math teachers employed by MSP Program schools between 2002 and 2004 relative to the percentage change in math teachers employed over this period by non-MSP Program schools.

## DATA

Data for studying the effect of the MSP in Texas come from five sources. Our analysis makes use of data from the MSP Program's MIS, the Texas State Board for Educator Certification (SBEC), the U.S. Department of Education's Common Core of Data, and labor market data from the Bureau of Economic Analysis and the U.S. Bureau of Labor Statistics. The primary use of the MIS data, collected from all MSP K-12 schools and IHEs, is to identify schools in Texas that were involved in the MSP Program as "participating" schools, and thus have the potential for teacher quantity and quality effects tied to MSP participation.

There are three Texas MSPs that we cover in this study, all of which have a strong math component as a part of their MSP proposal.<sup>14</sup> The three MSPs in Texas that we use are the Alliance for Improvement of Mathematics Skills PreK-16 (AIMS), the El Paso Math and Science Partnership, and the Texas Middle and Secondary Mathematics Project.

The AIMS MSP, a \$4 million targeted NSF award, unites Del Mar Community College and Texas A&M University-Kingsville with nine independent school districts in south Texas. These districts serve approximately 30,000 students, of whom 61% are minority, mostly Hispanic, students and 50% are economically disadvantaged students. The overarching goal of AIMS is to "prepare students in these partner districts for success in college-level mathematics courses by the time they graduate from high school."<sup>15</sup> One of the specific goals of AIMS is to

---

<sup>14</sup>A fourth Texas MSP, the Rice University Mathematics Leadership Institute, did not start until 2004, so schools from this MSP are not included in our MSP group.

<sup>15</sup>Information on AIMS is found at <http://www.delmar.edu/aims/Files/AIMS%20PR%20project%20summary.doc>

“conduct research on the extent to which the partnership increases . . . the number, diversity, and quality of math teachers” in the MSP schools.

The El Paso Math and Science Partnership is a \$29.3 million NSF award that includes three urban school districts that encompass El Paso; nine proximate rural school districts; the University of Texas at EL Paso; El Paso Community College; the Region 19 Education Service Center; and El Paso area civic, business, and community organizations and leaders. The first stated priority of this partnership is “more high-quality math . . . teachers.” The routes to meeting this goal are via “enhancing University of Texas at EL Paso’s Master of Arts in Teaching Mathematics program, . . . creating a pre-master’s program, providing tuition support for these programs, supporting alternative certification programs, and providing coaching to high school math and science teachers.”<sup>16</sup>

The Texas Middle and Secondary School Mathematics Project is funded by a \$3 million NSF award. This partnership combines 12 independent school districts with Stephen F. Austin State University and has a stated goal of “increasing the number of qualified and certified math teachers for grades 4–12.”<sup>17</sup>

Thus, an examination of the goals of the three Texas MSPs used in this study shows that each has an expectation that MSP involvement will lead to an increase in the number of math teachers they employ or at least an increase in the number of teachers available for employment. Of particular relevance for this study is that a close reading of the goals and priorities of these MSP partnerships makes it clear that their intention is an increase not simply in the number of math teachers, but rather an increase in the number of *certified* math teachers.

Data on the number of certified math teachers employed at given school over time, the information needed to construct the dependent variable in this study, come from the Texas SBEC. The SBEC data contain information on the number of in-field and out-of-field FTEs employed in every subject and on the number of uncertified teachers by subject in every public school in Texas, and for each of the years 2002, 2003, and 2004. We also use the SBEC data to control for the total number of math FTEs employed in school  $i$  in year  $t$ .

Variables in the  $X$  vector of Equation 1 that come from the Common Core of Data are

- the natural logarithm of the number of students in the high school
- the percentage of Black, Hispanic, Asian, and American Indian students in the high school
- the percentage of female students in the high school
- the percentage of students on free or reduced-price lunch in the high school

<sup>16</sup>Information on the El Paso MSP is found at <http://epcae.org/msp/msp.htm>.

<sup>17</sup>Information on the Texas Middle and Secondary School Mathematics Project is found at <http://www.faculty.sfasu.edu/kchilds/nsf2.html>.

- the percentage of students identified as immigrants

Variables in the  $Q$  vector of Equation 1 that come from the Common Core of Data are

- the total revenue per student in the district from all sources
- the percentage of total dollars in revenue at the district level that are raised locally
- the total district expenditures
- the average per FTE instructional staff salary in the district

Labor market variables in the  $Q$  vector are actually at the county, rather than the district level and include the average wage rate in the county in which school  $i$  is located (in 2002 constant dollars) and the average unemployment rate in that county for the year. This labor market information is included to capture the outside opportunity costs facing math teachers in district  $j$  in year  $t$ .

Table 1 gives information on the sample selection from the universe of Texas high schools into our analytic sample. We begin with 63 potential MSP schools in 35 different school districts and 1,594 non-MSP schools in the comparison group. Along with schools for which we could not construct control variables, we eliminate “alternative” schools such as schools at juvenile detention facilities, schools for which we lack data across all three years of the study, and schools that employed zero certified math teachers in any of the three years of the study.<sup>18</sup> This leaves us with 54 MSP schools in 34 school districts and 1,143 non-MSP schools. Of these 54 MSP schools, 37 schools in 19 different school districts satisfied the criteria we use to define MSP “participating” schools. Of these 37 schools, 35 were in the MSP Program for both 2003 and 2004, and two schools were in an MSP project for only 2004.

As discussed earlier, focusing on the number of math teachers at the school level could lead to a downward biased estimate of any positive MSP effect, as most employment decisions, particularly the number of teachers hired, are made at the district level. For example, it could be the case that participation in an MSP project pushes out the labor supply curve for a given school but that school is not able to act on this situation because of decisions made in the district central office. Thus, for this school the estimated MSP effect on the percentage change in certified math teachers employed would be zero. Because of this potential downward bias we fit some models using only MSP schools that are located in districts where *all* high schools participate in the MSP project. When we limit our sample to this group we have 34 MSP schools in 16 school districts.

---

<sup>18</sup>Because we are looking at the percentage change in the number of math teachers within a school, this construct makes sense only for schools that have at least one certified math teacher.

TABLE 1  
Sample Selection of Texas High Schools Into the Analytic Sample

	MSP Schools	Non-MSP Schools
Initial base year sample	63	1,594
Non-missing school- and district-level control variables	61	1,514
Non-missing labor market control variables	60	1,437
Not an “alternative” high school	60	1,252
Observed in all three years (2002, 2003, 2004)	60	1,250
Non-zero number of certified math teachers in all three years	54	1,143
Satisfied criterion to be an MSP “participating” school	37	NA
MSP school in 2003	35	NA
MSP school in 2004	37	NA
MSP school in a district with all MSP schools	34	NA

*Note.* MSP = Math and Science Partnership; NA = not applicable.

Summary statistics for our sample of Texas high schools for the base year of 2002 and for the last year, 2004, are presented in Table 2. The most obvious and important information from Table 2 is that the MSP participant schools are different from non-MSP high schools in Texas. The MSP schools in both 2002 and in 2004 were substantially larger than the non-MSP schools (i.e., they had more students on average), and hence they tended to employ more teachers, in both math and in all the other subject areas, than the non-MSP schools. The MSP schools also tended to hire slightly more noncertified or out-of-field math teachers. The MSP schools have higher percentages of black and Hispanic students and a higher percentage of students on free or reduced-price lunch.

At the district level, districts in which MSP schools are located have higher annual total revenue and expenditures, and a lower percentage of their revenues are raised locally. There is little difference in annual teacher salaries across the districts of MSP and non-MSP schools, but MSP schools tend to be located in counties with lower average annual salaries (across all occupations) and higher unemployment rates. These are all factors that could affect the supply of certified teachers, and thus we will control for these time-varying factors in our models.

RESULTS

Given the development and discussion of our analytic model, evaluations of the impact of the MSP participation on the supply of certified math teachers are based on the fixed effects model represented in Equation 1. In this section we present preliminary results from this model using data from Texas on high school math



TABLE 2

Summary Statistics (Means) for MSP and Non-MSP Texas High Schools in 2002 and 2004

	2002		2004	
	MSP	Non-MSP	MSP	Non-MSP
No. of schools	37	1,143	37	1,143
FTE certified math teachers	8.6 (5.3)	5.7 (5.4)	8.3 (5.2)	5.6 (5.3)
FTE certified teachers teaching math out of field	1.3 (1.1)	0.9 (1.3)	1.3 (1.4)	0.8 (1.1)
FTE noncertified teachers teaching math	0.9 (0.9)	0.7 (1.1)	1.1 (1.3)	0.6 (1.0)
FTE teachers in all other subjects	69.2 (38.7)	47.5 (43.5)	69.3 (39.2)	46.1 (42.2)
Total students	1,265 (791)	821 (828)	1,278 (817)	836 (895)
% students female	0.48	0.49	0.48	0.49
% White students	0.21	0.57	0.20	0.55
% Black students	0.04	0.11	0.04	0.11
% Hispanic students	0.73	0.30	0.75	0.32
% students other race/ethnicity	0.02	0.01	0.01	0.01
% students on free or reduced-price lunch	0.68	0.51	0.75	0.55
% immigrant students	0.05	0.04	0.08	0.05
Total annual district revenue (federal, state, and local in \$1,000)	201,758 (180,130)	143,058 (320,586)	225,342 (200,006)	162,698 (361,200)
Total annual district expenditures (in \$1,000)	197,496 (173,186)	154,773 (351,335)	231,222 (206,410)	169,994 (374,814)
Average teacher salaries in the district	43,168 (2708)	42,715 (4768)	45,724 (2996)	45,091 (5314)
% district revenues raised locally	0.30	0.44	0.29	0.46
Average annual salary in the county	26,522 (1863)	28,911 (7768)	28,143 (1919)	30,138 (7546)
Average county unemployment rate	0.068	0.050	0.082	0.066

*Note.* Standard deviations are in parentheses where appropriate. MSP = Math and Science Partnership; FTE = full-time equivalent.

teachers. We highlight the preliminary nature of the results because we have, at most, only two years of data when Texas MSP schools were involved in the project.

The discussion of our results begins with a reminder that, second to developing a model, a further goal of this study is to determine the extent to which our model might prove useful for future explorations of the impact of the MSP Program on teacher quantity when more appropriate data become available. Thus, we focus attention on the reasonableness of our point estimates, noting from the outset that none of our impact estimates are statistically significant.

We use three criteria for judging the "reasonableness" of the estimates. First, given the model, we would expect point estimates of the impact of MSP participation to be non-negative except as a result of sampling error. Second, we would

expect the impact from two years of MSP participation to be at least as large, if not larger, than the impact from one year of participation. Third, for reasons discussed earlier, we would expect estimates based on a sample that uses only those MSP schools that are in districts where *all* schools are in the MSP Program to be larger than estimates from a sample that employs all MSP schools regardless of MSP “saturation” within the district. With those guidelines for assessing the performance of our model, we turn to the estimates.

The first column of estimates in Table 3 is from a model that controls only for the log number of teachers, the average teacher salary in the district, and the log number of students in the school. The statistical interpretation of the estimates in the first column is that there is no evidence of an “MSP effect” on within-school changes in FTE certified math teachers. Were they statistically significant, the estimates of  $\delta_1$  and  $\delta_2$  would mean that on average MSP schools tended to hire about 3.7% more FTE certified math teachers between 2002 and 2003 and about 4.2% more between 2002 and 2004 than did non-MSP schools, controlling for the log total number of math teachers, average teacher salary in the district, the log number of students in the school, and school fixed effects.<sup>19</sup> Model 2 adds the time varying school-level and district- (or county) level covariates. With these additional controls, the point estimates fall by about 25% (2003) and 33% (2004) and remain statistically insignificant.

The third and fourth columns present estimates based on a sample that uses only those MSP schools that were in districts where all of the schools were MSP participants. As explained earlier, results that use schools in districts that are less than 100% saturated with the MSP Program across their high schools may be downwardly biased. The movement of the point estimates from columns 1 and 2 to columns 3 and 4 are supportive of this hypothesis as each estimate in the latter columns is larger than its companion estimate in columns 1 and 2.

Taken collectively, we read the estimates in Table 3 as supportive of what we would expect given our model. First, they are all non-negative as the model would suggest. Second, all sets of estimates indicate equal or larger second-year effects than first-year effects, as we would expect. Finally, as expected, we do see larger estimates of  $\delta_1$  and  $\delta_2$  when we move from using all MSP schools to using only those in MSP “saturated” districts (Model 1 vs. Model 3 and Model 2 vs. Model 4). We emphasize that we are not trying to draw any inferences from the Table 3 estimates. Rather, we are simply asking whether the patterns of point estimates are supportive or nonsupportive of the model.

To examine the extent to which different MSP projects may have differential effects, we estimate Model 4, our preferred model, using the three different MSP Programs. The estimates from these regressions are in Table 4.

<sup>19</sup>The exact interpretation is that the difference is  $e^\beta - 1$ , which is very close to  $\beta$  for the small values of  $\beta$  in most all of our results.

TABLE 3

Log FTE Regressions for Certified Math Teachers in Texas High Schools, 2002–2004, based on Equation 1

	Using All MSP Schools		Using Only Those MSP Schools In Districts With 100% MSP Participation	
	Model 1	Model 2	Model 3	Model 4
MSP*Y03 ( $\delta_1$ )	0.037 (0.087)	0.027 (0.088)	0.046 (0.096)	0.036 (0.097)
MSP*Y04 ( $\delta_2$ )	0.042 (0.092)	0.028 (0.091)	0.050 (0.100)	0.036 (0.099)
Y03	−0.004 (0.012)	−0.013 (0.027)	−0.004 (0.012)	−0.013 (0.027)
Y04	0.011 (0.014)	−0.015 (0.035)	0.011 (0.014)	−0.015 (0.036)
Log total no. of math teachers	0.253 (0.084)	0.250 (0.079)	0.253 (0.084)	0.250 (0.079)
Average teacher salary in the district	Yes	Yes	Yes	Yes
Log no. of students in the high school <sup>a</sup>	Yes	Yes	Yes	Yes
School-level controls, district revenue and expense controls, and county labor market controls	No	Yes	No	Yes
Total no. of schools	1,180	1,180	1,171	1,171
[No. of MSP schools]	[37]	[37]	[34]	[34]

*Note.* Standard errors are in parentheses. FTE = full-time equivalent; MSP = Math and Science Partnership.

The point estimates across the different projects offer some evidence of differential effects by project, though again, none of the estimated effects are statistically different from zero, and Chow tests would offer no evidence that the estimates across projects are different from each other. Nevertheless, we note that MSP Project 3 has a different pattern to the estimates than do the other two projects. This may completely be a function of sampling error, but we do note one factor that differentiates Project 3 from the other projects: Of the three projects used in our study, Project 3 is the only one that focuses on *both* math and science rather than only on math. To see the extent to which the estimates in Table 3 are driven by the schools in Project 3, we refit the models in Table 3 using only the schools in Projects 1 and 2. Those results are in Table 5.

The estimates in Table 5 are even more supportive of the model. Again, all of the point estimates are non-negative, all point estimates of  $\delta_2$  are, as expected, larger than the point estimates of  $\delta_1$ , and finally, all estimates of  $\delta_1$  and  $\delta_2$  are larger when only schools in MSP saturated districts are used. We emphasize again that these are only observational differences, and we would be unable to reject the null hypothesis of no difference between any of the comparisons discussed here.

TABLE 4  
Log FTE Regressions, by MSP Project, for Certified Math Teachers in Texas High Schools, 2002–2004, Based on Equation 1

	Project 1 <sup>a</sup>	Project 2 <sup>b</sup>	Project 3 <sup>c</sup>
MSP*Y03 ( $\delta_1$ )	0.094 (0.140)	−0.050 (0.053)	0.029 (0.119)
MSP*Y04 ( $\delta_2$ )	0.197 (0.162)	0.097 (0.064)	−0.004 (0.112)
Year dummies	Yes	Yes	Yes
Log total math teachers	Yes	Yes	Yes
Average teacher salary in the district	Yes	Yes	Yes
Log no. of students in the high school <sup>a</sup>	Yes	Yes	Yes
School-level controls, district revenue and expense controls, and county labor market controls	Yes	Yes	Yes
Total no. of schools	1,143	1,139	1,163
[No. of MSP schools]	[6]	[2]	[26]

*Note.* Standard errors are in parentheses. FTE = full-time equivalent; MSP = Math and Science Partnership.

<sup>a</sup>Alliance for the Improvement of Mathematics Skills, Del Mar College. <sup>b</sup>Texas Middle and Secondary Mathematics Project, Stephen F. Austin University. <sup>c</sup>El Paso Math and Science Partnership, University of Texas at El Paso.

TABLE 5  
Log FTE Regressions for Certified Math Teachers in Texas High Schools, 2002–2004, Based on Equation 1 and Excluding Project 3 Schools

	Using All MSP Schools		Using Only Those MSP Schools in Districts With 100% MSP Participation	
	Model 1	Model 2	Model 3	Model 4
MSP*Y03 ( $\delta_1$ )	0.033 (0.088)	0.022 (0.085)	0.072 (0.111)	0.058 (0.109)
MSP*Y04 ( $\delta_2$ )	0.133 (0.088)	0.104 (0.108)	0.200 (0.122)	0.171 (0.123)
Year dummies	Yes	Yes	Yes	Yes
Log Total no. of math teachers	Yes	Yes	Yes	Yes
Average teacher salary in the district	Yes	Yes	Yes	Yes
Log no. of students in the high school <sup>a</sup>	Yes	Yes	Yes	Yes
School-level controls, district revenue and expense controls, and county labor market controls	No	Yes	No	Yes
Total no. of schools	1,154	1,154	1,145	1,145
[No. of MSP schools]	[11]	[11]	[8]	[8]

*Note.* Standard errors are in parentheses. FTE = full-time equivalent; MSP = Math and Science Partnership.



## DISCUSSION

In this article we model the supply and demand of certified teachers under (a) salary constraints and (b) the ability to fill vacancies with uncertified or out-of-field teachers. An estimating equation based on the supply and demand analytics is used to estimate the impact of MSP participation on the supply of certified math teachers available to a school or school district. The value of the model of supply and demand is that it provides the rationale for inferring shifts in underlying supply by examining changes in the observed hiring mix of certified and uncertified teachers.

The first application of the model as a tool for learning about possible MSP Program impacts on the quantity and quality of math teachers uses data from Texas. These data are the best currently available for studying this difficult question, but they are not perfect. The primary shortcomings of the Texas data used in this article are the short time horizon of the available data and the small sample of MSP “participating” schools. At the most, the MSP schools in the data had been in the program for only two years, and some had only been in the program for one year. Given the mechanisms through which participation in the MSP Program might shift the supply curve of certified math teachers, it could well take several years for any effects to appear. Also, given another year of MSP involvement, it is likely that additional MSP partner schools in Texas would meet our requirements for being classified as “participating” schools, increasing the number of “treatment” schools in the sample.

One way of gauging the reasonableness of our point estimates is to explore the predicted impact based on our, admittedly, nonsignificant estimates. For this exercise we consider the effect size associated with the point estimate when we only use Projects 1 and 2. The estimate of  $\delta_2$  in Model 4 of Table 5 is 0.171 ( $p = .16$ ). If this were the true impact of MSP participation, this effect would translate into the mean MSP school hiring an additional one and a half FTE certified math teacher between 2002 and 2004 ( $[e^{0.171} - 1] \times 8.0 = 1.5$ , where 8.0 is the mean number of math teachers in MSP schools in the base year of 2002). Thus, even if our results were statistically significant, it would likely be in the eye of the (policy) beholder whether these were effect sizes of policy significance. Whether and the extent to which the estimated impact might be larger in subsequent years awaits further data collection.

Another issue to consider in interpreting our results is that we are estimating only the “MSP effect” on the labor supply curves facing individual schools or districts as a result of *their* participation in the MSP Program. As stated earlier, this could occur via attracting certified teachers from the reserve pool of nonteaching, certified teachers or via attracting newly certified math teachers into the district. Notably, we are not able to directly estimate any effect that the MSP Program might have on the extra production of certified math teachers that might be available to

all schools, MSP and non-MSP participants. That is, we can say nothing directly about MSP effects that occur on the IHE side of the projects. Projects that attempt to estimate the extent to which the MSP Program increases the production of newly certified math and science teachers are a logical next direction for research. However, such work awaits the identification and construction of appropriate data containing information across pre- and postprogram years on the numbers of newly certified math and science teachers for a sample of MSP and non-MSP IHEs.

### ACKNOWLEDGMENT

This study is one in a series of substudies for the Math and Science Partnership Program Evaluation (MSP-PE) conducted for the National Science Foundation's MSP Program. The MSP-PE is conducted under Contract No. EHR-0456995. Since 2007, Bernice Anderson, Ed.D., Senior Advisor for Evaluation, Directorate for Education and Human Resources, has served as the National Science Foundation Program Officer. The authors, from Brown University, are John H. Tyler and Svetla Vitanova.

The MSP-PE is led by COSMOS Corporation in current partnership with George Mason University (GMU) and Brown University. Robert K. Yin (COSMOS) serves as Principal Investigator and Jennifer Scherer (COSMOS) serves as one of three Co-Principal Investigators. Additional Co-Principal Investigators and their collaborating institutions (including discipline departments and math centers) are Patricia Moyer-Packenham (USU, formerly GMU) and Kenneth Wong (Brown).

### REFERENCES

- Brewer, D. (1996). Career paths and quit decisions: Evidence from teaching. *Journal of Labor Economics*, 14(2), 313–339.
- Broughman, S. P., & Rollefson, M. R. (2000). *Teacher supply in the United States, sources of newly hired teachers in public and private schools: 1987–88 to 1993–94*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Darling-Hammond, L., Holtzman, D.J., Gatlin, S. J., & Heilig, J.V. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and Teacher effectiveness. *Education Policy Analysis Archives*, 13(42).
- Dolton, P. (1990). The economics of UK teacher supply: The graduate's decision. *Economic Journal*, 100(4), 91–104.
- Dolton, P., & van der Klaauw, W. (1995). Leaving teaching in the UK: A duration analysis. *Economic Journal*, 105, 431–444.
- Dolton, P., & van der Klaauw, W. (1999). The turnover of UK teachers: A competing risk explanation. *Review of Economic Studies*, 81(3), 543–549.
- Galchus, K. E. (1994). An analysis of the factors affecting the supply and demand for teacher quality. *Journal of Economics and Finance*, 18(2), 165–178.

- Glazerman, S., Mayer, D., & Decker, P. (2006). Alternative routes to teaching: The impacts of Teach for America on student achievement and other outcomes. *Journal of Policy Analysis and Management*, 25(1), 75–96.
- Goldhaber, D. D., & Brewer, D. (2000). Does teacher certification matter? High school certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22, 122–145.
- Gritz, M. R., & Theobald, N. D. (1996). The effects of school district spending priorities on length of stay in teaching. *Journal of Human Resources*, 31(3), 477–512.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2), 371–406.
- Hill, H. C. (2007). Mathematical knowledge of middle school teachers: Implications for the No Child Left Behind policy initiative. *Educational Evaluation and Policy Analysis*, 29(2), 95–114.
- Ingersoll, R. M. (2000). *Turnover among mathematics and science teachers in the U.S.* Washington, DC: National Commission on Mathematics and Science Teaching for the 21st Century.
- Ingersoll, R. M. (2003). Is there a shortage among mathematics and science teachers? *Science Educator*, 12(1), 1–9.
- Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A Descriptive analysis. *Educational Evaluation and Policy Analysis*, 24(1), 37–62.
- Manski, C. F. (1987). Academic ability, earnings, and the decision to become a teacher: Evidence from the National Longitudinal Study of the High School Class of 1972. In D. A. Wise (Ed.), *Public sector payrolls*. (pp. 291–312) Chicago: University of Chicago Press.
- Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, 13, 125–145.
- Mont, D., & Rees, D. (1996). The influence of classroom characteristics on high school teacher turnover. *Economic Inquiry*, 34(1), 152–167.
- Murnane, R. J., & Olsen, R. J. (1989). The effects of salaries and opportunity costs on duration in teaching: Evidence from Michigan. *The Review of Economics and Statistics*, 71(2), 347–352.
- Murnane, R. J., & Olsen, R. J. (1990). The effects of salaries and opportunity costs on length of stay in teaching: Evidence from North Carolina. *Journal of Human Resources*, 25, 106–124.
- Murnane, R. J., Singer, J. D., Willett, J. B., Kemple, J. J., & Olsen, R. J. (1991). *Who Will Teach? Policies That Matter*. Cambridge, MA: Harvard University Press.
- National Center for Education Statistics. (2002). *Qualifications of the public school teacher workforce*. Washington, DC: U.S. Department of Education.
- No Child Left Behind Act of 2001, Pub L. No. 107–110, 115 Stat. 1425 (2002).
- Rockoff, J. E., Kane, T. J., & Staiger, D. O. (2006). *What does certification tell us about teacher effectiveness? Evidence from New York City* (NBER Working Paper No. 12155). Cambridge, MA: National Bureau of Economic Research.
- Rowan, B., Chiang, F., & Miller, R. J. (1997). Using research on employees' performance to study the effects of teachers on students' achievement. *Sociology of Education*, 70, 256–284.
- Stinebrickner, T. R. (1998). An empirical investigation of teacher attrition. *Economics of Education Review*, 17(2), 127–136.
- Stinebrickner, T. R. (2001). A dynamic model of teacher labor supply. *Journal of Labor Economics*, 19(1), 196–230.
- Theobald, N. D., & Gritz, M. R. (1996). The effects of school district spending priorities on the exit paths of beginning teachers leaving the district. *Economics of Education Review*, 15(1), 11–22.
- van der Klaauw, W. (1996). *Expectations and career decisions: An analysis of teaching careers using expectations data*. New York: New York University.
- Zabalza, A. (1979). The determinants of teacher supply. *Review of Economic Studies*, 46(1), 131–147.

# The Assessment of Mathematics and Science Teacher Quality

Patricia S. Moyer-Packenham  
*Utah State University*

Johnna J. Bolyard  
*West Virginia University*

Anastasia Kitsantas and Hana Oh  
*George Mason University*

The purpose of this study was to examine the types of instruments being used to document mathematics and science teacher quality characteristics in 48 nationally funded mathematics and science education awards. Each of the 48 projects operationalized teacher quality and determined how to assess it. The main research questions examined the instruments awardees used to gather data on mathematics and science teacher quality, and the main characteristics of teachers examined by awardees. Results showed that awardees most frequently used surveys or questionnaires to assess characteristics of mathematics and science teacher quality. The most common teacher characteristics examined by awardees' included teacher behaviors, practices, and beliefs, followed by the assessment of subject and pedagogical knowledge, and the documentation of mathematics and science teachers' certification. A few new instruments were under development and in use to assess characteristics of teacher quality. Detailed information on the development and psychometric properties of the instruments used for these examinations was not available from the reports. Because awardees were at different stages in their funded activities and data collection efforts were ongoing at the time of this analysis, this study offers a preliminary and formative review of the use of assessments to document mathematics and science teacher quality characteristics among these awards.

In recent years, educators, researchers, and policymakers have sought to identify the characteristics of a highly qualified teacher (No Child Left Behind [NCLB],



2002). This goal presents a challenge because the literature on teacher quality is extensive and examines a wide range of empirical studies on teacher characteristics assumed to reflect teacher quality (Darling-Hammond, 2000; Darling-Hammond & Youngs, 2002; Rice, 2003; Wilson & Floden, 2003; Wilson, Floden, & Ferrini-Mundy, 2001). The goal is of particular importance to the mathematics and science education community where reports of international comparisons show that student performance in the United States is less than desirable in these subject areas (Hiebert et al., 2003). Student performance is often attributed to the quality, or lack thereof, of K-12 mathematics and science teaching. Although there is agreement that teacher quality is important, there is great variability in operationalizing the construct and even more variability in assessing it (Rice, 2003). Therefore, operationalizing and assessing quality, specifically in terms of mathematics and science teaching, is also yet to be clarified. This leads us to question, What have researchers learned about assessing mathematics and science teacher quality?

Current reform efforts have brought increased funding for national initiatives focusing on the quality of teachers in mathematics and science (see, e.g., <http://www.ed.gov/> or <http://nsf.gov/>). This funding has resulted in some of the most cutting edge research on mathematics and science teacher quality in funded awards throughout the country, including the National Science Foundation's Math and Science Partnership (NSF MSP) Program. The NSF states the following as goals of the MSP Program:

MSP serves students and educators by emphasizing strong partnerships that tackle local needs and build grassroots support to:

- Enhance schools' capacity to provide challenging curricula for all students and encourage more students to succeed in advanced courses in mathematics and the sciences;
- Increase the number, quality and diversity of mathematics and science teachers, especially in underserved areas;
- Engage and support scientists, mathematicians, and engineers at local universities and local industries to work with K-12 educators and students;
- Contribute to a greater understanding of how students effectively learn mathematics and science and how teacher preparation and professional development can be improved; and
- Promote institutional and organizational change in education systems—from kindergarten through graduate school—to sustain partnerships' promising practices and policies. (NSF, 2007)

The study presented here was designed to examine one aspect within these goals, namely, the instruments used by the MSP awards as part of their efforts toward documenting mathematics and science teacher quality (Item 2). In the

2005 NSF Committee of Visitor's review of the MSP Program, in the section of the report focusing on "Results: Outputs and Outcomes of NSF Investments," the Committee of Visitor's review indicated,

Processes for measuring growth in teacher content knowledge and effectiveness are less well-developed, but NSF should pay attention to pre- and post-testing of teachers, to classroom observation, and in general to ensuring that across projects the growth of teacher knowledge can be measured. (NSF, 2005, p. 17)

Our study is an effort to respond to this review by initially examining the types of instruments used by awardees in the MSP Program to gather data on characteristics of mathematics and science teacher quality. Our investigation focused on three areas: (a) the *characteristics* of mathematics and science teacher quality being assessed; in other words, how mathematics and science teacher quality was defined and operationalized by awardees in the MSP Program; (b) the *instrumentation* being used by awardees for teacher assessment; and (c) the *psychometric properties* of the instruments.

In the following sections, we describe the literature that led to the assignment of categories of instruments, describe instruments used to assess mathematics and science teacher quality by awardees in the MSP Program, and review the teacher quality characteristics the awardees examine. Because awardees were at different stages in their funded activities and data collection efforts were ongoing at the time of this analysis, this study offers a preliminary and formative review of the use of instruments to document mathematics and science teacher quality characteristics among these awards.

## WHAT TEACHER QUALITY CHARACTERISTICS ARE EXAMINED IN RESEARCH?

There are six characteristics commonly identified by researchers in studies examining the quality of mathematics and science teachers (Bolyard & Moyer-Packenham, 2008). These characteristics include teacher behaviors, practices, and beliefs; subject knowledge; pedagogical knowledge; experience; certification status; and general ability. Among these characteristics are variables gathered through assessment measures (i.e., responses to test items or teaching performance during an observation) and nonassessment measures (i.e., highest degree obtained or number of years of teaching experience; American Statistical Association, 2007). A definition of teacher quality is sometimes defended by the relationship that research has found between a teacher variable and some other variable, often student achievement. As we present some of the relevant research findings, it is important to keep in mind the controversy involved in such a definition. Teachers

are associated with high or low student achievement test scores even when they are not in control of the characteristics of the students assigned to their classes, and they are not in control of other events that happen to their classrooms that are unpredictable.

*Teachers' behaviors, practices, and beliefs* provide important information about mathematics and science teacher quality. This aspect of teacher quality is usually the subject of studies using observational methods or self-report data. For example, in one observational study researchers found that 15% of observed mathematics and science lessons were categorized as high quality, whereas 27% and 59% were labeled medium and low quality, respectively (Hiebert et al., 2003; Weiss & Pasley, 2004; Weiss, Pasley, Smith, Banilower, & Heck, 2003). Some observational studies show associations between practices of high school science teachers and better classroom discipline (Druva & Anderson, 1983) and kindergarten teachers' instructional practices and student gains in mathematics (Guarino, Hamilton, Lockwood, & Rathbun, 2006). Further results indicate that teachers often decide how to teach content and those decisions are influenced by teachers' beliefs. For example, Staub and Stern (2002) found that elementary students of teachers who held more constructivist beliefs did better on word problem tests than students whose teachers used a more direct-instruction approach. Other research indicates a positive relationship between teachers' reported use of standards-based instruction and student achievement (Hamilton et al., 2003).

*Subject knowledge* is a highly valued characteristic of mathematics and science teachers and refers to the teacher's knowledge of mathematics and science content. Reviews of research indicate links between teachers' subject preparation and effectiveness, although these results are not always clear (Darling-Hammond, 2000; Darling-Hammond & Youngs, 2002; Rice, 2003; Wilson & Floden, 2003; Wilson et al., 2001). Results of studies examining the relationship between teachers holding subject specific degrees and student achievement vary, although mathematics results are generally positive (Chaney, 1995; Goldhaber & Brewer, 1997a, 2000; Rowan, Chiang, & Miller, 1997). Similarly, studies measuring teachers' subject knowledge using undergraduate or graduate coursework in the subject generally show a positive relationship with students' mathematics achievement (Chaney, 1995; Monk, 1994; Monk & King, 1994). Effects of subject matter coursework in science are often dependent upon the area of science studied (i.e., physical, earth, or life sciences; Chaney, 1995; Druva & Anderson, 1983; Monk & King, 1994). The data suggest a generally positive relationship between subject-specific mathematics and science coursework and student achievement. Some authors describe the intersection of subject-specific knowledge and pedagogy as *pedagogical content knowledge* (Shulman, 1986) or *mathematical knowledge for teaching* (Hill & Ball, 2004); however, this aspect of teacher knowledge is yet to be widely utilized as a research variable in studies on teacher quality.

Teachers' *pedagogical knowledge*, or knowledge of teaching, is often researched as evidence of teacher quality using data such as degrees in education, educational coursework, and scores on exams measuring professional knowledge. Researchers have reported positive effects of teachers' pedagogical knowledge and preparation (Adams & Krockover, 1997; Ferguson & Womack, 1993; Grossman & Richert, 1988; Grossman et al., 2000; Guyton & Farokhi, 1987; Hansen & Feldhusen, 1994; Valli & Agostinelli, 1993). Generally, studies of teachers' pedagogical knowledge find positive relationships between education training and teacher effectiveness (Darling-Hammond, 2000). Courses taken in subject-specific pedagogy (i.e., mathematics education or science education) also appear to have a positive impact, particularly in mathematics at the middle and secondary level (Chaney, 1995; Monk, 1994). However, other results show little or no relationships (e.g., Rivkin, Hanushek, & Kain, 2005). Wilson and Floden (2003) noted that much of the research focuses on teacher education programs rather than on specific courses or experiences.

Some studies report positive relationships between teachers' years of *experience* and teacher effectiveness (Ehrenberg & Brewer, 1995; Ferguson, 1991; Fetler, 1999; Goldhaber & Brewer, 1997b; Greenwald, Hedges, & Laine, 1996; Hanushek, 1992, 1996). Reviewing studies examining the relationship between teacher experience and student achievement, Rice (2003) concluded a positive relationship between these variables, which was more pronounced during the first years of teaching at the elementary level and more constant at the secondary level. Although characteristics such as teacher experience and education are commonly identified as favorable characteristics in the teacher hiring process, some researchers argue that little of the variation in teacher quality is explained by these variables (Rivkin et al., 2005).

Mathematics and science teachers' *certification* status is used as an indicator of knowledge gained from teacher preparation (Darling-Hammond, 2000; Darling-Hammond & Youngs, 2002). Certification refers to the types of teaching certificates one holds (e.g., secondary mathematics certificate, algebra endorsement, or physical science certification). Researchers compare those who are fully certified and those who hold provisional or emergency certification (Darling-Hammond, 2000; Fetler, 1999; Goldhaber & Brewer, 2000). Several studies indicate an advantage in favor of fully certified teachers on measures of student achievement and teacher performance evaluations (Darling-Hammond, 2000; Fetler, 1999). Mathematics student achievement has been found to be positively associated with having a teacher who is certified in-field (Goldhaber & Brewer, 1997b; Hawk, Coble, & Swanson, 1985).

Teacher's *general intellectual abilities*, that is, those verbal and quantitative abilities that frequently qualify individuals for higher education, are also considered aspects of teacher quality. Studies generally report a positive relationship between measures of teachers' general and verbal abilities and their effectiveness



(Ehrenberg & Brewer, 1994; Ferguson, 1991; Ferguson & Ladd, 1996; Greenwald et al., 1996; Hanushek, 1971; Strauss & Sawyer, 1986). Other studies indicate mixed or negative results (Ehrenberg & Brewer, 1995; Hanushek, 1992; Murnane & Phillips, 1981).

In this section we have classified the characteristics that researchers of teacher quality have included in their studies. We now turn to the ways that these characteristics have been measured.

## WHAT INSTRUMENTS ARE USED TO MEASURE TEACHER QUALITY?

Although much of the literature on teacher quality focuses on characteristics of teachers, there is less focus on the instrumentation used to gather data on those characteristics. In many cases, proxies, or substitutes for teacher quality characteristics, are used to measure the mathematics and science teacher quality construct, prompting different interpretations of the results in these studies (Darling-Hammond & Youngs, 2002; Wilson & Floden, 2003; Wilson et al., 2001). Some proxies are a better representation of the teacher quality characteristic than others. For example, studies use teachers' college majors as evidence of pedagogical and subject knowledge. However, a college major does not illuminate specific knowledge gained through such training or account for variations in programs among colleges and universities. The use of certification status is also common (Darling-Hammond, 2000; Goe, 2002; Goldhaber & Brewer, 2000). Yet states set their own certification criteria, and therefore, the skills and knowledge represented by a teacher's certification varies from state to state. Another difficulty is that teacher quality researchers sometimes use several variables that are highly correlated with each other. For example, education levels are highly correlated with age, experience, and general ability, and certification is often correlated with educational training and subject knowledge background (Darling-Hammond & Youngs, 2002). Combined with variations in units of analysis and methodological approaches, researchers may obtain conflicting results based on the same teacher characteristics.

Common instruments used to gather data on teacher quality in mathematics and science include written surveys and questionnaires, behavioral observations, exams, interviews, portfolios, and archival records. Researchers use written *surveys and questionnaires* to gather information about teachers' classroom practices and beliefs about teaching and learning (Darling-Hammond, Chung, & Frelow, 2002). Some surveys gather information on beginning teachers' professional concerns and opinions about their preparation (Darling-Hammond et al., 2002; Houston, Marshall, & McDavid, 1993; Sandlin, Young, & Karge, 1992). Surveys are sometimes used to gather information about teachers' entry into the profession (Andrew,

1990; Andrew & Schwab, 1995; Darling-Hammond et al., 2002), their perceptions of teaching as a profession (Lutz & Hutton, 1989), and their intention to remain in the profession (Darling-Hammond et al., 2002). Other surveys collect background information on teachers to use as representations of teacher quality characteristics (i.e., number of graduate and undergraduate courses taken, undergraduate institution, certification status, and major; Andrew, 1990; Andrew & Schwab, 1995).

*Behavioral observations* are often used to gather information on teachers' pedagogical knowledge and instructional practices. Observation protocols gather information on teachers' classroom management and instructional skills (Sandlin et al., 1992) and look for evidence of the use of best practices (Hawk et al., 1985; Miller, McKenna, & McKenna, 1998). Some observation data are examined to determine relationships between teachers' preservice preparation and their practices, knowledge, and beliefs (Adams & Krockover, 1997; Ferguson & Womack, 1993; Grossman, 1989; Grossman & Richert, 1988; Grossman et al., 2000; Hansen & Feldhusen, 1994). Generally these studies involve small sample sizes and combine observational data with data gathered through other sources. Observations of teacher behaviors and classroom practices provide a rich source of data, and there are several studies that have examined teachers' practices on a large scale (see, e.g., Weiss et al., 2003).

Scores on *exams* have been used to measure teacher characteristics such as subject knowledge, pedagogical or professional knowledge, and general or verbal ability. Exams are of two types: those used to measure subject knowledge created specifically for a study, and standardized exams such as the National Teachers Examination Subject Area Specialty exams (Hawk & Schmidt, 1989; Rowan et al., 1997) and the Praxis Subject Area exams. Exams used to measure teachers' pedagogical or professional knowledge include state and national certification exams such as the National Teachers Examinations Test of Professional Knowledge exam (Hawk & Schmidt, 1989). Some researchers have developed exams designed to measure the mathematical knowledge that teachers use in their work, or *mathematical knowledge for teaching* (MKT) (see, e.g., Hill & Ball, 2004). Scores on college entrance exams, such as ACT and SAT, and tests of verbal aptitude or basic literacy, are often used to measure teachers' general or verbal ability (Ferguson, 1991; Ferguson & Ladd, 1996; Hanushek, 1992).

*Interview* protocols are used to gather information on characteristics such as teachers' pedagogical knowledge and beliefs on teaching and learning. Interview data are often examined to determine relationships between teachers' preservice preparation and their practices, knowledge, and beliefs. Interview protocols are commonly used in conjunction with other instruments such as observations and surveys (Adams & Krockover, 1997; Ferguson & Womack, 1993; Grossman, 1989; Grossman & Richert, 1988; Grossman et al., 2000; Hansen & Feldhusen, 1994).

*Portfolios* and other written documents are analyzed as evidence of teachers' pedagogical skills and knowledge (Guyton & Farokhi, 1987). For example, one

study analyzed classroom artifacts (lesson plans and other teaching documents) from 10 beginning teachers to determine impacts of teacher education (Grossman et al., 2000). To apply for National Board Certification, teachers create teaching portfolios that contain videotapes of their teaching, evidence of student learning products, and a detailed analysis of their teaching practices (National Board for Professional Teaching Standards, <http://www.nbpts.org>).

*Archival records* often contain background information on teachers including degree completion, college transcripts and grade point average, college entrance exam scores, scores on professional certification exams, certification status, and years of experience. Data on certification status, degree completion, and graduate and undergraduate courses taken are often used as evidence of teachers' pedagogical and/or subject matter preparation (Chaney, 1995; Darling-Hammond, Holtzman, Gatlin, & Heilig, 2005; Fetler, 1999; Laczko-Kerr & Berliner, 2002; Monk, 1994; Rowan et al., 1997). The information is often gathered in and accessed through state and national databases.

In this section we have reviewed a variety of instruments commonly used to gather data on the quality of individual teachers. At this point we turn our attention to the characteristics of teacher quality identified by awardees in the MSP Program and the instruments used by awardees to assess those characteristics. Our analysis focused on the following research questions: (a) What instrumentation is being used by awardees to assess teacher quality characteristics? Two subquestions emerged from this research question: Are the instruments locally or externally developed? What information is available regarding the psychometric properties of the instruments being used? The second research question was (b) What teacher characteristics are being assessed by the instruments? Subquestions included the following questions: How is subject knowledge (mathematics, science, and MKT) measured? In this case it was hypothesized that standard content tests would be used to assess subject knowledge. How is pedagogical knowledge measured? It was hypothesized that surveys and observations would be used to assess pedagogical knowledge. In a further analysis we examined similarities and differences among the awardees in terms of when they received their awards (i.e., Cohort I, II, and III awards, distributed to partnerships between 2002 and 2004).

## METHODS

### Data Sources

The data sources in this study came from funded partnerships in the NSF-MSP Program awarded between fiscal year (FY) 2002 and FY2004. The NSF describes the following four components that make up the MSP Program:

- Comprehensive Partnerships implement change across the K-12 continuum in mathematics, science, or both.
- Targeted Partnerships focus on improved student achievement in a narrower grade range or disciplinary focus in mathematics and/or science.
- Institute Partnerships develop mathematics and science teachers as school- and district-based intellectual leaders and master teachers.
- Research, Evaluation, and Technical Assistance (RETA) activities assist partnership awardees in the implementation and evaluation of their work (NSF, 2007).

Our study examined data from 48 awards in three of these categories including 12 Comprehensive Partnerships, 28 Targeted Partnerships, and 8 Institute Partnerships. RETA awards were not included in the analysis because of the nature and scope of their work in “assisting” the other award categories.

Each partnership is required to address the quality of the mathematics and science teaching force and to document its progress toward the teacher quality goals and benchmarks it has established. Awardees submit Annual and Evaluation Reports describing this progress. In this analysis, researchers reviewed 123 Annual and Evaluation Reports provided to the NSF, with the length of each report ranging from 29 to 707 pages. These reports, along with awardees Web sites, published papers, and presentations, were the secondary source documents for the analysis. Data reviewed for this article were obtained from documents available to researchers between January 2005 and February 2006.

## DEFINING INSTRUMENT AND TEACHER QUALITY CHARACTERISTICS CATEGORIES

Based on the review of research, we determined a set of categories for types of instruments and a set of categories for teacher quality characteristics. The following sections define each of these categories and describe how they were used in the analysis.

### Instrument Categories

To focus the scope of the analysis, researchers determined the following criteria for the instruments that would be included in the analysis. One criterion was that the instrument needed to gather data on *teacher* quality, and the analysis was confined to instruments used with teachers. Teachers were defined as those whose primary instructional responsibilities were in the classroom with students for at least 50% of a school day. There were a variety of instruments in use among the



awards that collected data on attributes of other school positions (i.e., principals, administrators, curriculum specialists). Researchers selected those instruments that collected data on *teachers* for inclusion.

Another criterion was that the instruments needed to be used to collect data on *individual characteristics* of teachers. Individual teacher characteristics identified in the research included teacher behaviors, practices, and beliefs; subject knowledge; pedagogical knowledge; experience; certification status; and general ability (Bolyard & Moyer-Packenham, 2008). Instruments that collected data on teacher quantity and diversity, such as numbers of participants in courses and demographics on teacher race and ethnicity, were beyond the scope of our analysis because they focused on characteristics of teachers as a group or population rather than on the quality of the individual teacher. In-depth examinations of teacher quantity and diversity are the focus of other investigations in the MSP Program Evaluation (Moyer-Packenham, Bolyard, Oh, Kridler, & Salkind, 2006; Moyer-Packenham, Parker, Bolyard, Kitsantas, & Huie, 2008; Tyler & Vitanova, 2007).

We used the definition of an *instrument* based on research compiled by Prus and Johnson (1994) for categorizing instruments. This categorization system included six types of instruments: (a) written surveys and questionnaires, (b) behavioral observations, (c) exams, (d) exit and other interviews, (e) portfolios, and (f) archival and other records. By using this system of categorization, we limited the scope of the analysis, thereby excluding some types of data that were collected by the awardees. For example, many teachers in the partnerships attended courses and workshops to improve their knowledge and practices. When the MSPs reported *offering* a course or *numbers of teachers taking* a course, we had no way of knowing what teacher characteristics were impacted and what types of instruments were used in the course, and therefore course participation was not captured in this analysis. However, when the awardees reported their use of exams, interviews, or any other instruments to document teacher characteristics during or following courses, these instruments were included in our analysis. This type of focused examination ensured that the teacher characteristics assessed were linked directly by the awardees themselves with the instruments used to document the characteristics.

In this section we provide specific detail on the instrument categories as they relate to the present study. A *survey or questionnaire* was a document where respondents replied to questions or comments in writing, often choosing from a given set of answers (Fraenkel & Wallen, 1993). *Behavioral observations* included instruments, such as protocols, which categorize teacher behaviors and performances in a natural setting such as a classroom (Miles & Huberman, 1984; Prus & Johnson, 1994; Schloss & Smith, 1999). *Exams* were those instruments administered to teacher-participants as part of the awardees' activities. This category often included instruments designed to test knowledge in one or more areas (i.e., mathematics or science; Fraenkel & Wallen, 1993; Prus & Johnson, 1994), through multiple-choice, short-answer, and essay formats, among others,

and included instruments developed externally (by an individual or group outside the award) and those developed locally (by an individual working within the award; Fraenkel & Wallen, 1993; Lopez, 1998; Prus & Johnson, 1994). *Exit and other interviews* required participants to discuss their perceptions, beliefs, knowledge, or experiences often in a face-to-face setting with questions posed by an interviewer (Fraenkel & Wallen, 1993; Prus & Johnson, 1994). *Portfolios* included collections of work samples and other documents produced and compiled by teachers over time, with the portfolios most often assessed using a rubric (Hart, 1994; Paulson, Paulson, & Meyer, 1991; Prus & Johnson, 1994). *Archival records* included documents regarding background and demographic information, or other file data (Prus & Johnson, 1994). In our study, this information was often provided by an existing file compiled by a university or school district and included data on teacher certification status, teacher exam scores, and years of experience. When the score from an exam was gathered from instruments not administered by awardees during their activities, and was obtained from external database sources, these were categorized as archival records rather than exams.

The final category, *unspecified*, was added and included instruments for which awardees did not provide sufficient information to determine the assessment being used. In these cases, awardees described assessing a particular teacher characteristic but did not specify the instrument used in the assessment. A cross-checking method was used to search Web sites, conference papers, and other available documents in an attempt to identify these instruments. The *unspecified* category was used when no additional information was available following this search. Researchers looked for examples of the instruments among the documents to determine the content of each instrument.

## Teacher Quality Characteristics Categories

Researchers used the following six categories for teacher quality characteristics identified in a literature review conducted by Bolyard and Moyer-Packenham (2008): (a) teacher behaviors, practices, and beliefs; (b) subject knowledge; (c) pedagogical knowledge; (d) experience; (e) certification status; and (f) general ability. An additional category, *unspecified*, was used when the specific teacher characteristic being assessed could not be determined based on the descriptive information provided by awardees. As in the case of instruments, a cross-checking method was used to search other available documents for this information. The following section describes each of the teacher quality characteristics categories as they relate to our study.

The category teacher behaviors, practices, and beliefs was further defined in two subcategories: *teacher behaviors and practices* and *teacher beliefs*. The teacher behaviors and practices category included what the teacher does in the classroom, for example, questioning strategies, instructional equity, classroom management,

and use of time. Teacher beliefs included beliefs about students' learning, such as beliefs about the way students learn content and beliefs about who can and cannot learn, and beliefs about content, such as teachers' views on the nature of the content and the best methods for teaching it. *Subject knowledge* refers to teachers' knowledge and understanding of concepts and topics related to specific content (Ferguson & Womack, 1993; Monk, 1994). In our study, subject knowledge refers to knowledge of mathematics and science content, and MKT. MKT, as defined by Hill and Ball (2004), is the specialized kind of content knowledge needed to teach mathematics and is part of the work of one of the RETA awards in the NSF-MSP Program. *Pedagogical knowledge* refers to knowledge of teaching and learning including knowledge of students' cognitive development, learning theories, and instructional approaches and strategies. *Experience* is defined as the total number of years a teacher has been teaching and/or the number of years a teacher has taught a specific grade level or subject area, although researchers note that experience can also include the substance, variety, and quality of one's experiences. *Certification* describes teachers' certification status (including whether they are emergency, provisionally, or fully certified), whether a teacher is certified in the field in which they are teaching, and whether teachers are *highly qualified* as defined by NCLB (2002). *General ability* refers to teachers' general intellectual academic and verbal abilities, often including evidence of language and mathematical proficiency.

## Procedures

Researchers conducted a preliminary analysis of the secondary source documents that focused on understanding the major themes of teacher quality, quantity, and diversity among the work of awardees prior to our study. This preliminary analysis indicated that the awardees in this program were engaged in a variety of activities designed to influence teacher quality, quantity, and diversity and that they had implemented numerous strategies for assessing their progress. The prior examination showed that the data collected on teacher quality primarily focused on changes in teachers' subject and pedagogical knowledge, their practices and beliefs, and their certification status. The data on teacher quantity focused on numbers of teachers participating in MSP activities and activities of the schools and universities associated with the MSP award. Data on teacher diversity focused on reporting race and ethnicity of participating teachers. Overall, the preliminary analysis showed that interventions identified by the awardees as influences on teacher quality, quantity, and diversity characteristics included new programs and coursework; professional development; teacher leadership; recruiting; preservice training; compensation; retention; linking science, technology, engineering, and mathematics (STEM) faculty with teachers; and induction. These results are discussed in another Math and Science Partnership Program Evaluation (MSP-PE) manuscript (Moyer-Packenham et al., 2006).



Building on this prior analysis, the team of researchers examined the secondary documents to locate information on the *instruments* in use by MSP awardees. The prior analysis indicated that there were numerous instruments in use among the awards. The challenge faced by researchers was in extracting this information because it was scattered in a variety of different locations throughout the reports. Researchers found that some awardees described numerous instruments, whereas others included little information about their instruments in the reports. In many cases, the actual instruments themselves were described by awardees but were not included in the reports.

Researchers used the previously described definitions for instrumentation and teacher quality characteristics to sort and classify the data, compiling the following information for each instrument: the name of the award using the instrument, the name of the instrument, the teacher quality characteristic assessed, type of instrument, source of the instrument (local or external to the award), information on psychometric properties, and instrument availability (whether a copy of the instrument was included in the reports or other documents). The research team scanned reports from the RETA awards of the MSP Program to cross check for instruments that might be under development in the RETAs and determine if these were in use by awardees. Instruments were categorized along two dimensions: the type of instrument used and the teacher characteristics assessed by the instrument. These categories were analyzed by examining relationships and using descriptive and chi-square tests.

## RESULTS

The first research question examined all instruments being used by awardees. A total of 282 instruments were identified across the 48 awards. Figure 1 shows the distribution of these instruments. This is an average of almost six instruments reported per award (5.88) at the time of our preliminary analysis. As Figure 1 shows, every awardee identified at least 1 instrument (three reported only 1), and some reported as many as 10, 12, or even 15 instruments.

As shown in Table 1, the majority of instruments used across the 48 awards were survey/questionnaires (37.9% of all the instruments identified) used by 87.5% of the awards. These were followed by exams (16.0%) used by 62.5% of awards, behavioral observations (14.2%) used by 62.5% of awards, exit and other interviews (10.6%) used by 50% of awards, portfolios (7.1%) used by 29.2% of awards, archival records (10.6%) used by 45.8% of awards, and finally instruments that were unspecified (3.5%) used by 16.7% of awards.

The 107 surveys and questionnaires that were identified collected data on a wide range of topics from several different teacher audiences. One example was a survey intended for teacher participants focusing on their perceptions of changes



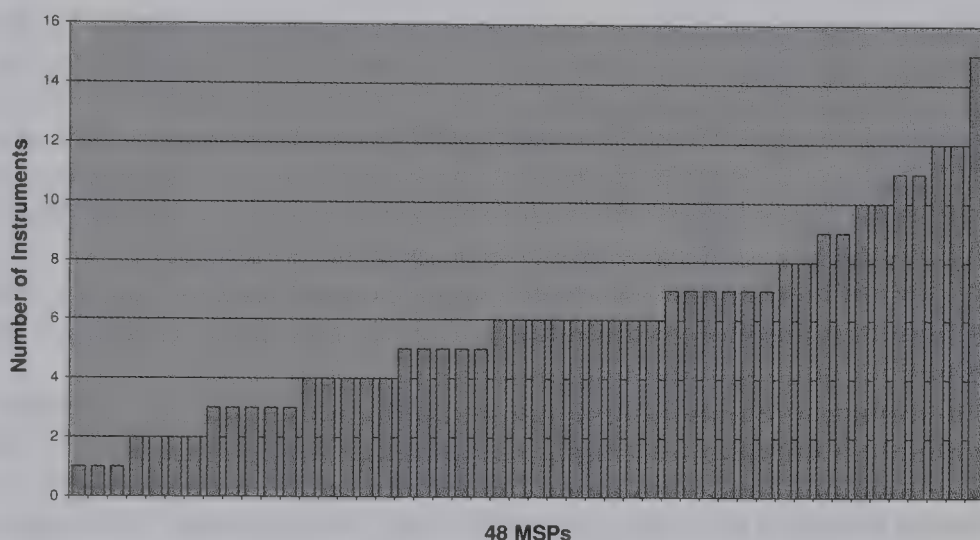


FIGURE 1 Distribution of instruments across Math and Science Partnership (MSP) awards.

in their knowledge, skills, and practices as a result of participation in an activity. In this example, the survey asked teachers about their perceptions of changes in their own knowledge rather than assessing their knowledge directly. One of the 45 exams assessed respondents' mathematical knowledge about precalculus concepts. Another exam was designed to measure growth in secondary teachers' knowledge of algebra and geometry. In the 40 behavioral observations, a variety of instruments asked observers to record information including demonstrated level of teachers' subject knowledge, tools and strategies employed, cognitive level of tasks, instructional equity, and lesson implementation. One example of the 30

TABLE 1  
Frequency and Percentage of Instruments Used Across the Awards

Instrument	Award Frequency <sup>a</sup>		Instrument Frequency <sup>b</sup>
	Not Used	Used	
Written surveys and questionnaires	6 (12.5%)	42 (87.5%)	107 (37.9%)
Exams	18 (37.5%)	30 (62.5%)	45 (16.0%)
Behavioral observations	18 (37.5%)	30 (62.5%)	40 (14.2%)
Exit and other interviews	24 (50.0%)	24 (50.0%)	30 (10.6%)
Portfolios	34 (70.8%)	14 (29.2%)	20 (7.1%)
Archival records	26 (54.2%)	22 (45.8%)	30 (10.6%)
Unspecified	40 (83.3%)	8 (16.7%)	10 (3.5%)

<sup>a</sup>  $N = 48$ ; <sup>b</sup>  $N = 282$

interviews used by awardees included an interview protocol designed to elicit information on changes in the teachers' own practices and their students' learning as a result of participation in the partnership. Among the 20 portfolios were those that analyzed teachers' writing in online logs to document successes, challenges, and concerns as teachers implemented award goals over time. Others focused on teachers' lesson plans to document changes in teachers' practices. The 30 archival records were documents that contained summative information about teacher licensure and certification status, years of experience, levels of education, grades and examination scores, and general ability measures (i.e., SAT or GRE scores).

In addition to examining type and frequency, researchers also determined whether the instruments were locally or externally developed, see Table 2. This examination was constrained to the documents available for analysis and was therefore limited in its scope. Locally developed instruments were those developed by awardees, whereas externally developed instruments were those developed by someone external to the award. This analysis revealed that the same number of surveys and questionnaires were locally developed and externally developed (30, or 28.0%), and 47 (43.9%) were not identified. Among the behavioral observations, 12 (30.0%) were locally developed, 17 (42.5%) were externally developed, and 11 (27.5%) were not identified. Exams tended to be externally developed (25, or 55.6%), whereas 9 (20.0%) were locally developed, and 11 (24.4%) were not identified. Most of the interview instruments were not identified (18, or 60.0%), 8 (26.7%) were locally developed and 4 (13.3%) were developed externally. In terms of the portfolios, 9 (45.0%) were locally developed, 2 (10.0%) were developed externally, and 9 (45.0%) were not identified. Finally, 1 (10%) of the unspecified documents was locally developed and 9 (90.0%) were not identified.

Next researchers examined the psychometric properties of the locally developed instruments. These results are also presented in Table 2. For 26 (86.7%) and 27 (90.0%) surveys and questionnaires there was no information reported about the validity and reliability, respectively. However, 4 (13.3%) reported validity information and 3 (10.0%) reported reliability. Similar patterns were observed for the behavioral observations instruments (1 of 12 reported validity and reliability) and exams (2 and 1 of 9 reported validity and reliability, respectively). No psychometric properties were reported for interviews or portfolios. Archival records were not included in this table because psychometric properties can not be established for this type of instrument.

Researchers conducted further investigations of the number of awards using exam instruments to measure types of subject-specific knowledge, including mathematics, science, and MKT; see Table 3. Among the 48 MSPs were awards that focused on mathematics only, science only, and a combination of mathematics and science. There were 40 awards that included mathematics, and 27 awards that included science. Of the 40 awards that included mathematics, 17 awards (42.5%)

TABLE 2  
Frequency and Percentage of Instrument Properties Across the Awards

Instrument	N	Locally Developed: Psychometric Properties						Externally Developed Frequency	Development Not Identified Frequency
		Frequency	Validity	Reliability Not Reported	Validity Reported	Reliability Reported	Frequency		
			Not Reported						
Written surveys and questionnaires	107	30 (28.0%)	26 (86.7%)		4 (13.3%)	27 (90.0%)	3 (10.0%)	30 (28.0%)	47 (43.9%)
Behavioral observations	40	12 (30.0%)	11 (91.7%)		1 (8.3%)	11 (91.7%)	1 (8.3%)	17 (42.5%)	11 (27.5%)
Exams	45	9 (20.0%)	7 (77.8%)		2 (22.2%)	8 (88.9%)	1 (11.1%)	25 (55.6%)	11 (24.4%)
Exit and other interviews	30	8 (26.7%)						4 (13.3%)	18 (60.0%)
Portfolios	20	9 (45.0%)						2 (10.0%)	9 (45.0%)
Unspecified	10	1 (10.0%)						0 (0.0%)	9 (90.0%)

*Note.* The 30 archival records are not included in this table.

TABLE 3  
Awardees' Use of External and Local Exams to Measure Types of Subject Knowledge

	Subject Knowledge Category							
	Exams Combined		Externally Developed Exams		Locally Developed Exams		Development Not Identified	
	Not Used	Used	Not Used	Used	Not Used	Used	Not Used	Used
Math content <sup>a</sup>	23 (57.5%)	17 (42.5%)	33 (82.5%)	7 (17.5%)	36 (90.0%)	4 (10.0%)	32 (80.0%)	8 (20.0%)
Science content <sup>b</sup>	14 (51.9%)	13 (48.1%)	21 (77.8%)	6 (22.2%)	23 (85.2%)	4 (14.8%)	23 (85.2%)	4 (14.8%)
MKT <sup>a,c</sup>	27 (67.5%)	13 (32.5%)	27 (67.5%)	13 (32.5%)	40 (100%)	0 (0.0%)	40 (100%)	0 (0.0%)

*Note.* Some awards use more than one type of exam with different sources of development; therefore numbers in the rows do not sum.

<sup>a</sup> *N* = 40 *Math-focused* Math and Science Partnerships.

<sup>b</sup> *N* = 27 *Science-focused* Math and Science Partnerships.

<sup>c</sup> Mathematical Knowledge for Teaching (MTK) as defined by Hill and Ball (2004).

used mathematics content exams to measure subject knowledge and 13 awards (32.5%) used the MKT instrument. Of the 27 awards that included science, 13 awards (48.1%) used science content exams to measure subject knowledge. Next we determined whether awards used exam instruments that were locally or externally developed. This analysis revealed that seven (17.5%) awards measuring mathematics content used exams that were externally developed, whereas four (10.0%) awards used locally developed exams, and eight (20.0%) used mathematics content exams whose development was not identified. In regards to exams measuring science content, six (22.2%) awards used exams that were externally developed, whereas four (14.8%) awards used locally developed exams, and four (14.8%) used exams where development was not identified. All of the awards that measured MKT (13 or 32.5%) used an exam that was developed external to the award.

The second research question examined the teacher characteristics being assessed by the instruments. Table 4 provides the frequencies of the teacher characteristics measured and not measured. Based on these results, 41 (85.4%) awards focused on assessing teacher behaviors, practices, and beliefs, with some awards focusing specifically on teachers' behaviors and practices only (37 or 77.1%), and others focusing on teachers' beliefs only (31 or 64.6%). Thirty-nine (81.3%) awards reported assessing subject knowledge, including 27 of 40 (67.5%) mathematics awards measuring mathematics knowledge, 18 of 27 (66.7%) science



TABLE 4  
Frequency and Percentage of Teacher Characteristics Examined by All Instruments Across the Awards

Teacher Characteristic	Frequency	
	Not Measured	Measured
Teacher behaviors, practices, and beliefs (combined) <sup>a</sup>	7 (14.6%)	41 (85.4%)
Teacher behaviors and practices	11 (22.9%)	37 (77.1%)
Teacher beliefs	17 (35.4%)	31 (64.6%)
Subject knowledge (combined) <sup>a</sup>	9 (18.8%)	39 (81.3%)
Math content <sup>b</sup>	13 (32.5%)	27 (67.5%)
Science content <sup>c</sup>	9 (33.3%)	18 (66.7%)
MKT <sup>b,d</sup>	27 (67.5%)	13 (32.5%)
Pedagogical knowledge	11 (22.9%)	37 (77.1%)
Certification	18 (37.5%)	30 (62.5%)
Experience	30 (62.5%)	18 (37.5%)
General ability	44 (91.7%)	4 (8.3%)
Unspecified	29 (60.4%)	19 (39.6%)

Note. *N* = 48.

<sup>a</sup>Combined totals reflect the number of awards measuring one or more characteristics in that category. <sup>b</sup>*N* = 40 *Math-focused* Math and Science Partnerships. <sup>c</sup>*N* = 27 *Science-focused* Math and Science Partnerships. <sup>d</sup>Mathematical Knowledge for Teaching (MKT) as defined by Hill and Ball (2004).

awards measuring science knowledge, and 13 of 40 (32.5%) mathematics awards measuring MKT. Pedagogical knowledge was assessed by 37 (77.1%) awards, whereas teacher certification was documented by 30 (62.5%) awards. Teacher experience and general ability were documented by 18 (37.5%) and four (8.3%) awards, respectively. Finally, 19 (39.6%) awards described instruments that measured teacher characteristics that could not be identified based on the descriptions in the reports.

Table 5 depicts the frequencies and percentages of the subquestions for research question two answering what teacher characteristics are being assessed. Regarding the first subquestion, how subject knowledge (combined) was assessed, nine awards used surveys and/or questionnaires, nine used behavioral observations, 30 used exams, four used interviews, five used portfolios, one used an archival record, and two awards did not specify. Pedagogical knowledge was assessed using surveys and/or questionnaires by 24 awards; 20 awards used behavioral observations, four used exams, 12 used interviews, seven used portfolios, two used archival records, and one award did not specify. Mathematics knowledge was assessed using surveys and/or questionnaires by five awards, whereas seven awards used behavioral observations, 17 used exams, three used interviews, four used portfolios, one used an archival record, and one did not specify. Science knowledge was assessed using

TABLE 5  
Awardees' Use of Instruments to Document Types of Teacher Knowledge

Teacher Knowledge Category	Instrument Uses											
	Surveys		Observations		Exams		Interviews		Portfolios		Archival Records	
	Not Used	Used	Not Used	Used	Not Used	Used	Not Used	Used	Not Used	Used	Not Used	Used
Subject knowledge (combined) <sup>a</sup>	39 (81.3%)	9 (18.8%)	39 (81.3%)	9 (18.8%)	18 (37.5%)	30 (62.5%)	44 (91.7%)	4 (8.3%)	43 (89.6%)	5 (10.4%)	47 (97.9%)	1 (2.1%)
Math content <sup>b</sup>	35 (87.5%)	5 (12.5%)	33 (82.5%)	7 (17.5%)	23 (57.5%)	17 (42.5%)	37 (92.5%)	3 (7.5%)	36 (90.0%)	4 (10.0%)	39 (97.5%)	1 (2.5%)
Science content <sup>c</sup>	22 (81.5%)	5 (18.5%)	24 (88.9%)	3 (11.1%)	14 (51.9%)	13 (48.1%)	26 (96.3%)	1 (3.7%)	25 (92.6%)	2 (7.4%)	26 (96.3%)	1 (3.7%)
MKT <sup>b,d</sup>	40 (100%)	0 (0.0%)	40 (100%)	0 (0.0%)	27 (67.5%)	13 (32.5%)	40 (100%)	0 (0.0%)	40 (100%)	0 (0.0%)	40 (100%)	0 (0.0%)
Pedagogical knowledge <sup>a</sup>	24 (50.0%)	24 (50.0%)	28 (58.3%)	20 (41.7%)	44 (91.7%)	4 (8.3%)	36 (75.0%)	12 (25.0%)	41 (85.4%)	7 (14.6%)	46 (95.8%)	2 (4.2%)

*Note.* Some awards use multiple instruments; therefore numbers in the rows do not sum.

<sup>a</sup>  $N = 48$  Math and Science Partnerships. <sup>b</sup>  $N = 40$  Math-focused Math and Science Partnerships. <sup>c</sup>  $N = 27$  Science-focused Math and Science Partnerships.

<sup>d</sup> Mathematical Knowledge for Teaching (MKT) as defined by Hill and Ball (2004).

surveys and/or questionnaires in five awards, whereas three awards used behavioral observations, 13 used exams, one used an interview, two used portfolios, one used an archival record, and one did not specify. Finally, MKT was measured using an exam in 13 of the 40 mathematics awards.

Chi-square tests were used to test the hypotheses that (a) standard content tests would be used to measure subject knowledge (mathematics, science, and MKT), rather than observations, surveys, portfolios, or interviews, whereas (b) surveys, observations, and interviews would be used to assess teacher's pedagogical knowledge rather than exams. Support for this hypothesis was found. First, in terms of mathematics knowledge, a significant  $\chi^2 (6, N = 146) = 12.80, p < .05$  was obtained, showing that exams were more often used to capture teacher content knowledge in mathematics. Similar results were revealed for science,  $\chi^2 (6, N = 146) = 15.01, p < .05$ , and MKT,  $g\chi^2 (6, N = 146) = 33.08, p < .001$ . Moreover, as hypothesized, awards used surveys and observations,  $\chi^2 (6, N = 146) = 90.00, p < .001$ , to assess teachers' pedagogical knowledge, which is significantly different from the way that subject knowledge was measured.

Finally, in regards to the last research question researchers examined the data for similarities and differences among the awards in the types of teacher characteristics examined and the number and type of instruments used by Cohort I, II, and III awards (awarded between FY2002 and FY 2004). The first subquestion focused on the types of teacher characteristics assessed by different cohorts of awards. Essentially, this examination showed that the awards in each cohort were using similar instruments to gathering data on the same teacher quality characteristics, and no overall significant differences emerged for teacher characteristics; see Table 6. At a descriptive level, frequencies showed that 90.9% of Cohort I assessed teachers' behaviors, practices, and beliefs, as compared with 85.7% of Cohort II and 75.0% of Cohort III. This trend was similar for the assessment of subject knowledge by the awards in Cohorts I (77.3%), II (71.4%), and III (50%). Although 68.2% of Cohort I and 75.0% of Cohort III awards assessed pedagogical knowledge, a larger portion of the Cohort II awards (92.9%) assessed this characteristic. The assessment of certification status (Cohort I, 63.6%; Cohort II, 64.3%; and Cohort III, 58.3%) and teacher experience (Cohort I, 36.4%; Cohort II, 42.9%; and Cohort III, 33.3%) were similar across the three cohorts. All cohorts focused less on collecting data on general ability (Cohort I, 0.0%; Cohort II, 14.3%; and Cohort III, 16.7%).

In regards to the second part of the final research question, no significant differences were detected among the frequency of instruments within each instrument category among the Cohort I, II, and III awards; see Table 7. Descriptively, more instruments were used in each instrument type in relation to the year that the MSP was awarded their funding (i.e. Cohort I, awarded 2002, 138 instruments; Cohort II, awarded 2003, 92 instruments; and Cohort III, awarded 2004, 53 instruments). There were also more documents available for analysis from the awards that were

TABLE 6  
Frequency and Percentage of Teacher Characteristics Examined by Cohort I, II, and III Awards

Teacher Characteristic	Cohort I (Awarded 2002) <sup>a</sup>		Cohort II (Awarded 2003) <sup>b</sup>		Cohort III (Awarded 2004) <sup>c</sup>	
	Not Measured	Measured	Not Measured	Measured	Not Measured	Measured
Teacher behaviors, practices, and beliefs	2 (9.1%)	20 (90.9%)	2 (14.3%)	12 (85.7%)	3 (25.0%)	9 (75.0%)
Subject knowledge	3 (13.6%)	19 (86.4%)	2 (14.3%)	12 (85.7%)	4 (33.3%)	8 (66.7%)
Pedagogical knowledge	7 (31.8%)	15 (68.2%)	1 (7.1%)	13 (92.9%)	3 (25.0%)	9 (75.0%)
Certification	8 (36.4%)	14 (63.6%)	5 (35.7%)	9 (64.3%)	5 (41.7%)	7 (58.3%)
Experience	14 (63.6%)	8 (36.4%)	8 (57.9%)	6 (42.9%)	8 (66.7%)	4 (33.3%)
General ability	22 (100%)	0 (0.0%)	12 (85.7%)	2 (14.3%)	10 (83.3%)	2 (16.7%)

<sup>a</sup>N = 22. <sup>b</sup>N = 14. <sup>c</sup>N = 12.

funded earlier, and these awards had more data collection activities accumulated over the years they had invested in their award. Therefore, the earlier the MSP was awarded, the more documents there were available for researchers to analyze, resulting in a larger number of instruments reported. However, when the proportions were compared for each instrument type, the three cohorts were all using instruments in similar proportions. These results indicate that, although the make-up of the three cohorts contained different types of partnerships, the types of instruments used and the teacher quality characteristics assessed were similar among the cohorts.

LIMITATIONS

Researchers acknowledge several limitations in our study. A major limitation was our exclusive use of secondary source documents to gather data about the instruments in use by these awardees. Because this was a preliminary analysis of the MSP-PE, researchers were constrained to the use of documents provided by the awardees to the funding agency through annual reports, evaluation reports, published papers, presentations, and project Web sites. This limited our data in several ways. First, awardees were not required to describe and include samples of their instruments and assessments or their psychometric properties in their reports to the funding agency. For this reason, the information on the instruments was reported voluntarily by awardees and is potentially an underrepresentation of the actual



TABLE 7  
Frequency and Percentage of Instruments Used by Cohort I, II, and III Awards

Instrument	Cohort I (Awarded 2002) <sup>a</sup>	Cohort II (Awarded 2003) <sup>b</sup>	Cohort III (Awarded 2004) <sup>c</sup>
Written surveys and questionnaires	50 (36.2%)	41 (45.1%)	16 (30.2%)
Behavioral observations	22 (15.9%)	10 (11.0%)	8 (15.1%)
Exams	18 (13.0%)	16 (17.6%)	11 (20.8%)
Exit and other interviews	15 (10.9%)	10 (11.0%)	5 (9.4%)
Portfolios	11 (8.0%)	4 (4.4%)	5 (9.4%)
Archival records	15 (10.9%)	8 (8.8%)	7 (13.2%)
Unspecified	7 (5.1%)	2 (2.2%)	1 (1.9%)

*Note.* *N*s indicate the number of instruments in each of the Cohort I, II, and III awards.

<sup>a</sup> *N* = 138. <sup>b</sup> *N* = 91. <sup>c</sup> *N* = 53.

amount of instruments in use. In addition, researchers were not able to interact with the awardees at the time of this analysis because the MSP-PE was in its early stages and had not yet gained permission to collect data directly from awardees. This prevented researchers from interviewing awardees to determine instruments in use that may not have been identified in the secondary source documents.

Another limitation is the element of time. While researchers were gathering and analyzing data from the secondary documents, awardees were going on with their work and developing and using additional instruments to collect data on characteristics of teacher quality. For example, one RETA has designed a knowledge assessment for middle school science teachers, focusing on *Force and Motion*, *Plate Tectonics*, and *Flow of Matter and Energy in Living Systems* (Smith, 2007). This assessment has an inventory of 1,170 items covering K-12 physical science and earth science content standards. Although this assessment was not identified by any of the awardees at the time of our investigation, it may be in use by awardees at the time our study is in print. Therefore, the results reported here represent a previous point in time along the continuum of the ongoing work of these awards. Additional analyses of the instrumentation among awardees will be enhanced by the MSP-PE's ability to gather new data directly from awardees in the future.

Although our study was limited in its scope, we believe that it serves a useful purpose in providing an initial examination of the instrumentation in use among awardees in the MSP Program, thereby providing a formative assessment and impetus for comprehensive reporting on instrumentation for assessing characteristics

of teacher quality. The identification of the instruments in use by awardees in this study is also a useful first step toward determining how to design further examinations of the growth of teacher content knowledge, which was an important goal put forth in the Committee of Visitor's review of the MSP Program (NSF, 2005).

## DISCUSSION

The results of our study show the instrumentation used by awardees to assess teacher quality characteristics in a national mathematics and science program. The findings illustrate teacher characteristics of most importance to awardees and the instruments used to gather data on those characteristics. Several key findings emerged from the analysis.

### What the Results Reveal About the Assessment of Teacher Quality Characteristics

These results reveal that awardees in this program are engaged in the assessment of teacher quality using a variety of different types of instruments to document the growth of several teacher characteristics. Although much of the pure research in the general domain of teacher quality uses characteristics such as years of experience, general ability, and certification status as representations of teacher quality, awardees in our study were more likely to assess (a) teachers' behaviors, practices, and beliefs; (b) subject knowledge; and (c) pedagogical knowledge (85.4%, 81.3%, and 77.1% of awards, respectively). In the context of this awards program these results are not surprising. These are characteristics for which the awardees have identified specific goals for improvement as part of their work. The awards are funded based on a set of project-specific goals and plans for demonstrating and assessing progress toward those goals. It makes sense that awardees would focus assessments of teacher quality on subject matter knowledge; pedagogical knowledge; and behaviors, practices, and belief, because these are characteristics of teachers over which awardees' work may have some influence.

Exams were used most often to assess subject knowledge, and surveys and observations were used most often to assess pedagogical knowledge. The use of exams to assess subject knowledge was true for all three types of subject knowledge (mathematics, science, and MKT). The use of exams is a common and preferred method for assessing subject knowledge in academic settings, including schools and universities. Because each of these awards is a partnership among schools and universities, with discipline faculty involved in the teacher knowledge development work of the award, using exams is viewed as a practical and objective measure for this characteristic. More than half of the awards in our study used exams that were developed externally. Reasons for this may be that exams are

more available to the awardees as resources from external sources than other types of instruments. In addition, the development of exams is a complex and time-intensive process that involves a variety of psychometric processes to validate the instruments. The use of surveys or observations to assess pedagogical knowledge was consistent among many of the awards in the program. In some cases, a combination of surveys, observations, and exams was used to gather data on teacher characteristics for partnership activities. Teacher quality is a complex construct, and it was not uncommon for awardees to utilize various instruments to collect data on different teacher characteristics in the hopes that these data could be triangulated to illuminate teacher change. The use of various instruments reveals that awardees are aware of the complexity inherent in documenting teacher growth and that they are attempting to focus on that growth as it relates to teachers' participation in partnership activities.

### The Quality of the Instruments

Almost every award used surveys and questionnaires, with almost one third of these developed locally by awardees. However, the awards in this analysis were not required to provide comprehensive information about the instruments in use at the time of this review, and therefore much of the information on the psychometric properties of the locally developed instruments was unknown. In contrast, 28% of the instruments were identified as externally developed, which means that the potential for these instruments to have psychometric properties is promising. An additional 37% of the instruments in use did not have their development identified, and perhaps some of these have available psychometric properties as well. Because the development of so many of the instruments was not identified, and because many were not available for direct review, researchers could not reach any general conclusions about the quality of these instruments.

In future research and development work that includes the creation and use of instruments to assess teacher quality, reporting psychometric properties of the instrumentation will be informative to researchers and educators. When conclusions are reached in any assessment of teacher quality characteristics without reporting sufficient information about the instrumentation used in the assessment, careful attention must be given to the trustworthiness of the results. Inclusion of this information in publications by the awardees will be a necessary part of the interpretation of any findings. In the case of these data, previously discussed limitations prevented researchers in our study from determining if the instruments did not have psychometric properties or if this information was simply not included in the secondary source documents because it was not required.

The limited amount of information being widely distributed on the instruments currently in use by the awardees is a drawback to others engaged in mathematics and science teacher quality work. Researchers in our study recommend that

awardees organize and expand the collection of MSP instruments available. Although one of the RETA Awards (<http://www.addingvalue.org>) currently lists several resources for instrumentation, and the MSP Toolbox/Materials section of the MSP Net Web site (<http://hub.mspnet.org/index.cfm/join>) lists some instruments, this resource could be expanded more broadly. In addition to awardees posting their instruments for shared access, the site could be a place to post standards for the selection of high-quality instruments in an effort to support awardees and enhance the quality of data gathered in the MSP Program. Standards for selecting instruments should include basic questions such as, What criteria were used to select the instruments? How do we know that this instrument is gathering evidence that will help us to determine whether or not we have reached our project's benchmarks and goals? These are good practices to adopt in evidence-based designs and beneficial when instruments are discussed and shared with the broader research community.

### The Development of Instruments That Fill Needed Niches

An important idea that emerged from these findings for the general field of teacher education research is that there are a limited number of instruments available that effectively measure mathematics and science teacher quality characteristics. As NCLB set the goals for teacher accountability, and educators sought to achieve "Highly Qualified" teacher status, greater focus was placed on assessing the quality of mathematics and science teachers. National and international comparisons in mathematics and science painted a less than favorable picture of the quality of America's mathematics and science teaching force. As a result, benchmarks were set to ensure that every mathematics and science classroom would have a highly qualified teacher. A need developed for assessments of teacher characteristics that better reflected teacher quality. As part of this process, important questions have emerged. For example, What instruments are specific to measuring the quality of mathematics and science teachers? Are there measures of mathematics and science teacher quality that can be tied to student outcomes? Is it possible to develop instruments to assess the multidimensional characteristics needed to effectively teach mathematics and science?

Prior research has indicated that there are gaps in the instrumentation available to measure types of teacher knowledge. Developing and testing these instruments is time-consuming and expensive work. But there is evidence among these awards that instruments are under development and in use by awardees in this program. For example, the MKT assessment, which was not developed solely with funds from this program, is the result of ongoing research from a variety of funding sources including an NSF-MSP RETA award (Hill & Ball, 2004; Hill, Rowan, & Ball, 2005). This instrument filled a needed niche for assessing subject and teaching knowledge for mathematics at the elementary level, whereas previous assessments focused on measuring mathematics subject knowledge alone. Because the MKT



instrument is being used and tested in settings across 13 of the awards, it provides an opportunity for its developers to gather data on its use across a variety of mathematics teaching and learning environments.

Although the goals of awardees were not specifically focused on the development of new instruments, almost one fourth of the instruments identified in this analysis were reported as developed locally (69 instruments) by the awardees themselves. About 10% of these had also reported some psychometric properties at the time of this analysis. Among these instruments are assessments that have the potential to fill needed niches for collecting data on other teacher quality characteristics. These newly developed instruments appear in a number of different categories (surveys, observations, exams, interviews, and portfolios) and may be particularly useful to schools and universities because they were developed by awardees in the program and used in applied settings. New instruments that assess mathematics and science teacher quality at the end of preservice training at the university, for the purpose of hiring mathematics and science teachers for K-12 school positions, or to identify areas of needed in-service training for teachers, would benefit the field of education and the assessment of mathematics and science teacher quality.

## CONCLUSION

At the beginning of this research our team posed the following question: What have researchers learned about assessing mathematics and science teacher quality? The results of our study shed some light on the answers to this question. Our findings indicate that there are a variety of instruments in use for assessing characteristics of mathematics and science teacher quality, including exams, surveys, observations, and interviews. The characteristics of mathematics and science teachers most commonly assessed among these awards included teacher behaviors, practices, and beliefs; subject knowledge; and pedagogical knowledge, which the research indicates are teacher characteristics commonly associated with student achievement outcomes. There are also a number of instruments that have been developed and are under development for assessing characteristics of mathematics and science teachers. These developing instruments may fill gaps that currently exist in instrumentation, providing researchers and educators with better ways to assess mathematics and science teacher quality.

## ACKNOWLEDGMENT

This study is one in a series of substudies for the Math and Science Partnership Program Evaluation (MSP-PE) conducted for the National Science Foundation's

MSP Program. The MSP-PE is conducted under Contract No. EHR-0456995. Since 2007, Bernice Anderson, Ed.D., Senior Advisor for Evaluation, Directorate for Education and Human Resources, has served as the National Science Foundation Program Officer. The authors are Patricia S. Moyer-Packenham, Utah State University (formerly of George Mason University), Johnna J. Bolyard, West Virginia University, Anastasia Kitsantas and Hana Oh, George Mason University.

The MSP-PE is led by COSMOS Corporation in current partnership with George Mason University (GMU) and Brown University. Robert K. Yin (COSMOS) serves as Principal Investigator and Jennifer Scherer (COSMOS) serves as one of three Co-Principal Investigators. Additional Co-Principal Investigators and their collaborating institutions (including discipline departments and math centers) are Patricia Moyer-Packenham (USU, formerly GMU) and Kenneth Wong (Brown). Any opinions, findings, conclusions, and recommendations expressed in this article are those of authors and do not necessarily reflect the views of the National Science Foundation.

## REFERENCES

- Adams, P. E., & Krockover, G. H. (1997). Beginning science teacher cognition and its origin in the preservice secondary science teacher program. *Journal of Research in Science Teaching*, 34, 633–653.
- American Statistical Association. (2007). *Using statistics effectively in mathematics education research*. Retrieved February 25, 2007, from [http://www.amstat.org/research\\_grants/pdfs/SMERReport.pdf](http://www.amstat.org/research_grants/pdfs/SMERReport.pdf)
- Andrew, M. D. (1990). Differences between graduates of 4-year and 5-year teacher preparation programs. *Journal of Teacher Education*, 41(2), 45–51.
- Andrew, M. D., & Schwab, R. L. (1995). Has reform in teacher education influenced teacher performance? An outcome assessment of graduates of an eleven-university consortium. *Action in Teacher Education*, 17(3), 43–53.
- Bolyard, J. J., & Moyer-Packenham, P. S. (2008/this issue). A review of the literature on mathematics and science teacher quality. *Peabody Journal of Education*, 83(4), 509–535.
- Chaney, B. (1995). *Student outcomes and the professional preparation of eighth-grad teachers in science and mathematics* (National Science Foundation Rep. No. RED-9255255). Rockville, MD: Westat. (ERIC Document Reproduction Service No. ED 389530)
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives*, 8(1). Retrieved February 2005 from <http://epaa.asu.edu/epaa/v8n1>
- Darling-Hammond, L., Chung, R., & Frelow, F. (2002). Variation in teacher preparation: How well do different pathways prepare teachers to teach? *Journal of Teacher Education*, 53, 286–302.
- Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). *Does Teacher Preparation Matter? Evidence about Teacher Certification, Teach for America, and Teacher Effectiveness*. Retrieved August 16, 2005, from Stanford University, Center for Teacher Education and School Reform Web site: <http://schoolredesign.net/srn/news/certification.html>
- Darling-Hammond, L., & Youngs, P. (2002). Defining “highly qualified teachers”: What does “scientifically-based research” actually tell us? *Educational Researcher*, 31(9), 13–25.

- Druva, C. A., & Anderson, R. D. (1983). Science teacher characteristics by teacher behavior and by student outcome: A meta-analysis of research. *Journal of Research in Science Teaching*, 20, 467-479.
- Ehrenberg, R. G., & Brewer, D. J. (1994). Do school and teacher characteristics matter? Evidence from High School and Beyond. *Economics of Education Review*, 13(1), 1-17.
- Ehrenberg, R. G., & Brewer, D. J. (1995). Did teachers' verbal ability and race matter in the 1960s? Coleman revisited. *Economics of Education Review*, 14(1), 1-21.
- Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation*, 28, 465-498.
- Ferguson, R. F., & Ladd, H. F. (1996). How and why money matters: An analysis of Alabama schools. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 265-298). Washington, DC: Brookings Institution.
- Ferguson, P., & Womack, S. T. (1993). The impact of subject matter and education coursework on teaching performance. *Journal of Teacher Education*, 44, 55-63.
- Fetler, M. (1999). High school staff characteristics and mathematics test results. *Education Policy Analysis Archives*, 7(9). Retrieved December 15, 2006, from <http://epaa.asu.edu/epaa/v7n9>
- Fraenkel, J. R., & Wallen, N. E. (1993). *How to design and evaluate research in education* (3rd ed.). New York: McGraw-Hill.
- Goe, L. (2002). Legislating equity: The distribution of emergency permit teachers in California. *Education Policy Analysis Archives*, 10(42). Retrieved June 2, 2005, from <http://epaa.asu.edu/v10n42/>
- Goldhaber, D. D., & Brewer, D. J. (1997a). Evaluating the effect of teacher degree level on educational performance. In W. J. Fowler (Ed.), *Developments in school finance, 1996* (pp. 197-210). Washington, DC: National Center for Education Statistics, U.S. Department of Education.
- Goldhaber, D. D., & Brewer, D. J. (1997b). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *The Journal of Human Resources*, 32, 505-523. Retrieved May 25, 2005, from JSTOR database.
- Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22(2), 129-146.
- Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, 66, 361-396.
- Grossman, P. L. (1989). Learning to teach without teacher education. *Teachers College Record*, 91, 191-207.
- Grossman, P. L., & Richert, A. E. (1988). Unacknowledged knowledge growth: A re-examination of the effects of teacher education. *Teaching and Teacher Education*, 4, 53-62.
- Grossman, P. L., Valencia, S. W., Evans, K., Thompson, C., Martin, S., & Place, N. (2000). Transitions into teaching: Learning to teach writing in teacher education and beyond. *Journal of Literacy Research*, 32, 631-662.
- Guarino, C. M., Hamilton, L. S., Lockwood, J. R., & Rathburn, A. H. (2006). *Teacher qualifications, instructional practices, and reading and mathematics gains of kindergarteners. Research and development report* (NCES No. 2006-031). Washington, DC: National Center for Education Statistics.
- Guyton, E., & Farokhi, E. (1987). Relationships among academic performance, basic skills, subject matter knowledge, and teaching skills of teacher education graduates. *Journal of Teacher Education*, 38(5), 37-42.
- Hamilton, L. S., McCaffrey, D. F., Stecher, B. M., Klein, S. P., Robyn, A., & Bugliari, D. (2003). Studying large-scale reforms of instructional practice: An example from mathematics and science. *Educational Evaluation and Policy Analysis*, 25, 1-29.
- Hansen, J. B., & Feldhusen, J. F. (1994). Comparison of trained and untrained teachers of gifted students. *Gifted Child Quarterly*, 38, 115-121.

- Hanushek, E. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *The American Economic Review*, 61, 280–288.
- Hanushek, E. (1992). The trade-off between child quantity and quality. *The Journal of Political Economy*, 100(1), 84–117. Retrieved January 20, 2006, from JSTOR database.
- Hanushek, E. (1996). A more complete picture of school resource policies. *Review of Educational Research*, 66, 397–409.
- Hart, D. (1994). *Authentic assessment: A handbook for educators*. Menlo Park, CA: Addison-Wesley.
- Hawk, P. P., Coble, C. R., & Swanson, M. (1985). Certification: It does matter. *Journal of Teacher Education*, 36(3), 13–15.
- Hawk, P. P., & Schmidt, M. W. (1989). Teacher preparation: A comparison of Traditional and alternative programs. *Journal of Teacher Education*, 40(5), 53–58.
- Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., et al. (2003). *Teaching Mathematics in Seven Countries: Results from the TIMSS 1999 Video Study* (No. NCES 2003–013 Revised). Washington DC: U.S. Department of Education, National Center for Education Statistics.
- Hill, H. C., & Ball, D. L. (2004). Learning mathematics for teaching: Results from California's mathematics professional development institutes. *Journal for Research in Mathematics Education*, 35(5), 330–351.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2), 371–406.
- Houston, W. R., Marshall, F., & McDavid, T. (1993). Problems of traditionally prepared and alternatively certified first-year teachers. *Education and Urban Society*, 26(1), 78–89.
- Laczko-Kerr, I., & Berliner, D. C. (2002). The effectiveness of "Teach for America" and other under-certified teachers on student academic achievement: A case of harmful public policy. *Educational Policy Analysis Archives*, 10(37). Retrieved March 4, 2005, from <http://epaa.asu.edu/epaa/v10n37/>
- Lopez, C. L. (1998). Assessment of student learning. *Liberal Education*, 84(3), 36–43.
- Lutz, F. W., & Hutton, J. B. (1989). Alternative teacher certification: Its policy implications for classroom personnel practice. *Educational Evaluation and Policy Analysis*, 11, 237–254.
- Miles, M. B., & Huberman, A. M. (1984). *Qualitative data analysis*. Beverly Hills: Sage.
- Miller, J. W., McKenna, M. C., & McKenna, B. A. (1998). A comparison of alternatively and traditionally prepared teachers. *Journal of Teacher Education*, 49, 165–176.
- Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, 13(2), 125–145.
- Moyer-Packenham, P. S., Bolyard, J. J., Oh, H., Kridler, P., & Salkind, G. (2006). Representations of teacher quality, quantity, and diversity in a national mathematics and science program. *Journal of Educational Research & Policy Studies*, 6(2), 1–40.
- Moyer-Packenham, P. S., Parker, J., Bolyard, J. J., Kitsantas, A., & Huie, F. (2008, May). *Examining strategies that promote teacher diversity in mathematics and science*. Research Paper Presentation, Twelfth Consultation of the International Consortium for Research in Science and Mathematics Education (ICRSME), Quito, Ecuador.
- Monk, D. H., & King, J. A. (1994). Multilevel teacher resource effects in pupil performance in secondary mathematics and science: The case of teacher subject matter preparation. In R. G. Ehrenberg (Ed.), *Choices and consequences: Contemporary policy issues in education* (pp. 29–58). Ithaca, NY: ILR Press.
- Murnane, R. J., & Phillips, B. R. (1981). What do effective teachers of inner-city children have in common? *Social Science Research*, 10, 83–100.
- National Science Foundation. (2005). *Memorandum: Committee of visitors report for the math and science partnership program*. Retrieved December 20, 2006, from <http://www.nsf.gov/od/oia/activities/cov/ehr/2005/MSPcov.pdf>



- National Science Foundation. (2007). *MSP goals, structure and composition*. Retrieved February 21, 2007, from [http://www.nsf.gov/ehf/MSP/nsf05069\\_3.jsp](http://www.nsf.gov/ehf/MSP/nsf05069_3.jsp)
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).
- Paulson, L. F., Paulson, P. R., & Meyer, C. (1991). What makes a portfolio a portfolio? *Educational Leadership*, 48(5), 60-63.
- Prus, J., & Johnson, R. (1994). Assessment and testing myths and realities. *New Directions for Community Colleges*, 88, 69-83.
- Rice, J. K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes*. Washington, DC: Economic Policy Institute.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458.
- Rowan, B., Chiang, F., & Miller, R. J. (1997). Using research on employees' performance to the study effects of teachers on students' achievement. *Sociology of Education*, 70(4), 256-284.
- Sandlin, R. A., Young, B. L., & Karge, B. D. (1992). Regularly and alternatively credentialed beginning teachers: Comparison and contrast of their development. *Action in Teacher Education*, 14(4), 16-23.
- Schloss, P. J., & Smith, M. A. (1999). *Conducting research*. Upper Saddle River, NJ: Prentice-Hall.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4-14.
- Smith, P. (2007). *Assessing teacher learning about science teaching*. Retrieved February 26, 2007, from <http://atl.mspnet.org/>
- Staub, F. C., & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains: Quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology*, 94, 344-355.
- Strauss, R. P., & Sawyer, E. A. (1986). Some new evidence on teacher and student competencies. *Economics of Education Review*, 5, 41-48.
- Tyler, J., & Vitanova, S. (2008/this issue). Does MSP participation increase the supply of math teachers? Developing and testing an analytic model. *Peabody Journal of Education*, 83(4), 536-561.
- Valli, L., & Agostinelli, A. (1993). Teaching before and after professional preparation: The story of a high school mathematics teacher. *Journal of Teacher Education*, 44, 107-114.
- Weiss, I. R., & Pasley, J. D. (2004). What is high-quality instruction? *Educational Leadership*, 61(5), 24-28.
- Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. J. (2003). *Looking inside the classroom: A study of K-12 mathematics and science education in the United States*. Chapel Hill, NC: Horizon Research.
- Wilson, S., & Floden, R. E. (2003). *Creating effective teachers: Concise Answers for Hard Questions. An Addendum to the Report "Teacher preparation research: Current knowledge, gaps, and recommendations."* Education Commission of the States. (ERIC Document Reproduction Service No. ED 476366)
- Wilson, S. M., Floden, R., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations. A research report prepared for the U.S. Department of Education*. Seattle: Center for the Study of Teaching and Policy, University of Washington. Retrieved March 2005 from <http://www.ctpweb.org/>

# Mathematics Curriculum Systems: Models for Analysis of Curricular Innovation and Development

Margret A. Hjalmarson  
*George Mason University*

Defining challenging curriculum first requires an examination of what is meant by curriculum. This discussion of challenging curriculum is motivated by the evaluation of the National Science Foundation's Math and Science Partnership Program. Standards frameworks, textbooks, software, and pedagogy are some aspects of curricula. The level of challenge of a curriculum is a locally defined, qualitative characteristic that depends on the curriculum system. The structure of a curriculum system is proposed to investigate the purposes, representations, and conceptual systems inherent in models of curriculum that are part of mathematics teaching and learning initiatives. Three types of models are proposed: content focused, pedagogically focused, and learner centered. The models draw on examples from the Math and Science Partnership portfolio and from other areas of the literature on mathematics curriculum.

The Latin origin of curriculum is *currere*, meaning “to run,” which connects both to the curriculum as a course of study and to the content that students should learn in a given class. Running also implies motion. Curriculum runs over time and moves forward. With this forward movement comes inevitable change with different groups of students, different teachers, and different schools. One fundamental question is, What is meant by “the curriculum”? More specifically, what is being investigated for effectiveness or impact on student learning? Curriculum development for textbooks may include the design of tangible materials for students, supplementary materials for instructors, and software design. Curriculum development for districts and schools may include the design of guidelines, frameworks, and standards to aid alignment of curriculum, teaching, and assessment.

Curriculum may also include how teachers select and implement materials for teaching. In short, “curriculum” functions more as a system of interactive components at different levels of the educational system. This article examines the nature and design of curriculum systems in the context of the National Science Foundation’s Math and Science Partnership (NSF MSP) Program. The examination is set in the context of the MSP Program Evaluation (MSP-PE). The NSF did not specify any specific curriculum or type of curriculum as preferred in the grant solicitation so projects have responded with a range of interpretations of “challenging curriculum.”

My discussion of curriculum systems has two goals: first, to examine the operation of curriculum systems, and second, to present models of curriculum systems as analytic and descriptive tools for comparing curriculum system initiatives. The first section describes definitions of curriculum and provides more detail about the construct of a curriculum system as an integrated network of activities. In the second part, I present three types of models that can be used to categorize and analyze curriculum systems. The underlying structure for each model includes a conceptual system with a representation developed for a purpose.

Because the nature of curricula is complex and involves multiple components, I refer throughout this article to “curriculum systems” to include the aspects of mathematics teaching and learning that relate to curriculum. I adopt this vocabulary to distinguish from curriculum as it refers to textbook materials or to frameworks (e.g., standards documents) as both of those are aspects of the system under alignment and investigation. A curriculum system includes the tangible materials used in classrooms with students, assessments of student learning, frameworks and standards to guide instruction, and the content included in any of the products mentioned.

## DEFINING CURRICULUM

Burkhardt, Fraser, and Ridgway (1990) described curriculum as part of a system that includes students in classrooms taught by teachers in schools. In part, the definition depends on the audience (e.g., designer, teacher, principal, parent, or teacher). There are tangible books and documents to point to and say “there’s the curriculum.” But the issue is more complex because, as with any object, a curriculum carries interpretations, expectations, and cultural values. The curriculum is not only the text in the textbook or the items in the standards but also a systemic instructional endeavor including pedagogical interaction between teachers and students using tangible materials. Clements’s (2002) definition of the curriculum as “an instructional blueprint and set of materials for guiding students’ acquisition of certain culturally-valued concepts, procedures, intellectual dispositions, and ways of reasoning” (p. 601) includes the social constructs that surround the

tangible objects that make up the curriculum in the eyes of various educational participants. Clements (2002) and Burkhardt et al. (1990) described six characterizations of curriculum—ideal, available, adopted, implemented, achieved, and tested—in their discussions of the links between curriculum and research. The classification proceeds from what the developers intend (the ideal curriculum) to what teachers may use (available and adopted) to what teachers do use (implemented) to learning at the student level (achieved and tested). MacNab's (2000) definition has a similar characterization of the three types of curriculum as the intended, the implemented, and the experienced. All of these authors describe the curriculum system as an entity open for interpretation by the users (either students or teachers) rather than a fixed entity situated only in tangible artifacts. Any curriculum system is then situated within a local context where the curriculum is ultimately evaluated. Clements (2007) provided a detailed framework for developing "research-based curriculum"; linking curriculum development and scientific research methods; and utilizing many stages of development, theory building, and refinement of materials.

At a broad level, MacNab (2000) discussed the intended, the implemented, and the experienced curriculum. The intended curriculum is what the designers planned for the curriculum to accomplish. The implemented curriculum is how the teacher uses the curriculum in the classroom. The experienced curriculum is how the students understand, interpret, and experience the curriculum. The experienced curriculum could be impacted by students' prior knowledge, experiences, and other student-level attributes. In the ideal situation, curriculum should be experienced and implemented as intended. Modifications and selections are made by the teacher to fit the curriculum within other parameters and constraints in the classroom (e.g., time, administrative expectations, scope, and sequence). Teachers make decisions about the content of instruction and need to be selective about the topics chosen from the textbook (Porter, 2002) so the enacted curriculum is more complex than what is in the textbook. The three phases of the curriculum system included multiple artifacts and multiple interpretations. The system is also dynamic as modifications are made at each stage.

At the first stage, the ideal or intended curriculum system is designed. Potential designers include curriculum experts, teachers, school districts, software developers, and other relevant experts. The ideal curriculum system is represented in artifacts such as standards documents, textbook materials, and software. In the second stage of the curriculum system (implemented/enacted), the artifacts are implemented in the system by the teacher (or teachers). The interpretation of the artifacts as implemented is critical for analyzing effectiveness because the "challenge" of a task is not only within the task itself but also in how it is used. For instance, Stein, Grover, and Henningsen (1996) found that a challenging mathematical task can become a nonchallenging mathematical task under certain teaching conditions. Under their definition of challenging, tasks needed to include opportunities for



higher order thinking about mathematics. They found that some teachers reduced the cognitive difficulty of the tasks by scaffolding. Artifacts in the implemented phase include lesson plans, classroom observations, and other objects the teacher may design within the curriculum system (e.g., assessment guides, homework assignments). At this phase, the teacher's beliefs, mathematical knowledge, and pedagogical knowledge enter the system. The enacted curriculum has been investigated in large-scale studies such as the Trends in International Mathematics and Science Study (Hiebert et al., 2003; Schmidt et al., 2001) but remains a complex aspect of curriculum research. A significant question regarding the implemented or enacted curriculum system is its match to the ideal curriculum system.

In the final stage of the system, there is the students' experience of a curriculum system. The artifacts include their mathematics assessments, classroom observations, and achievement data. Investigations into student experience go beyond the classroom and include other factors (e.g., socioeconomic status, gender, cultural factors). The students' experience is critical to understanding whether the ideal curriculum system meets its ideals, how the teacher has implemented the curriculum, and the effects on student learning. By the time the ideal curriculum system reaches the experienced level, it has been through two levels of interpretation (at least). However, the system captures the relevant participants and the types of artifacts that could represent their interpretations of the curriculum system and the content. Focusing on a system at the level of what is experienced by students helps expand beyond the materials used to the environment where they are used and other cultural, social, pedagogical, or environmental factors that may impact how a curriculum system is used in the classroom.

Related to MacNab's (2000) framework and the more detailed frameworks (e.g., Clements, 2002) is the question of balancing the local with what is generalizable. The challenge in the MSP-PE for mathematics curriculum is that there are multiple local curriculum initiatives occurring with multiple operational definitions of what is "good" for the particular local system. In addition, components of the curriculum system are continually reinterpreted in local contexts (e.g., as districts develop frameworks based on state standards, as teachers use materials with students). Observational studies of local contexts are useful (e.g., Henningsen & Stein, 1997) but challenging for a large-scale program evaluation such as MSP-PE. Because there are a large number of initiatives occurring simultaneously and because many grants are developing products such as learning units or curriculum frameworks, I focus next on the function of representations and artifacts of curriculum systems.

## Representations and Products of Curriculum Systems

To analyze the effectiveness or the level of challenge of a curriculum, the representations and artifacts of that system must be sources of information about what is valued by the system and how effectiveness is measured. "Products" is used

here to describe the objects included as part of an analysis of curriculum systems. They could include but are not limited to textbook materials, instructional materials, software, standards documents, and frameworks. At this point, there is a distinction between the artifacts of curriculum systems and the pedagogical methods even though there is interaction between the two. Such products are a representation of the goals, assumptions, and conceptual foundations behind the curriculum systems design. As with any representation, any individual artifact does not represent the whole system. Sets of artifacts more accurately portray the designers' intentions for the curriculum system.

The measurement of attributes of the systems is related to the representational artifacts. For instance, an effective curriculum, a user-friendly curriculum, or other adjectives to describe curriculum systems are measurable. What is notable about such characteristics is that they do not exist outside the system of use. For instance, curriculum systems that are user friendly for some students and teachers may not be so for other groups of students and teachers. For example, students who speak only Spanish will probably not find math activities written in English "user friendly." The context-dependent nature of qualitative characteristics of representations is critical to considerations such as whether or not a curriculum system "works" or is "effective." The qualitative characteristics are situated within the representations of the curriculum system. The representation impacts how the curriculum system is interpreted, implemented, and experienced. Another characteristic of representations is the potential for change in the systems. For example, curricula do not exist in isolation. Rather, they respond to and impact the systems where they are used. Curricula may change how teachers see their students or understand mathematics (Harris, Marcus, McLaren, & Fey, 2001; Lloyd, 2002; Middleton, 1999).

The representational nature of curriculum artifacts is emphasized here because the assessment and evaluation of curriculum systems is not only an assessment or evaluation of the materials themselves but also of the theories and intentions behind the curriculum system design (Clements, 2007). As with any representation, there is an interactive, reflexive relationship between the creator (or creators) of the representation and the interpretation of the artifact. As an example, *Number Power* is a *Standards*-based curriculum. When first implemented, teachers found the materials difficult to use. However, teachers' difficulties decreased with the next year of use (Battistich, Alldredge, & Tsuchida, 2003). The representation itself (e.g., the teacher's guides, the student materials) had not changed, but the interpretation of the representation changed over time, as did teachers' experience with using the curriculum.

The focus on the representational nature of the products also underscores their existence within a larger system (e.g., classroom, school, district) where multiple stakeholders interpret the materials in different ways. As an example, in a study of preservice teachers' views of textbooks, there was a mismatch between their assumptions about how the textbook should help students learn and the views of

teaching and learning represented by *Connected Mathematics* (Hjalmarson, 2005). Some preservice teachers expected the textbook to be a reference for students to use to find definitions or algorithms, however *Connected Mathematics* is not and does not aspire to be a reference book. As documented in other studies, teachers' views of their role impacts their use of the materials and students' learning (Lloyd, 1999, 2002; McCaffrey et al., 2001; Mokros, 2003; Pligge, Kent, & Spence, 2000). For instance, didactic use of *Investigations* was less effective than use of the materials as intended by the designers (Mokros, 2003). However, as teachers' knowledge develops, their interpretations of materials representing the curriculum system (i.e., the representation of a view on teaching and learning) may change.

The curriculum system represents not only how students should learn but what content is valued in the context. As a representation of content, curriculum system designers place different value on different content within the system. Topics may be emphasized or not depending on the purposes of the design. As Mullis, Martin, Gonzalez, and Chrostowski (2004) found in the TIMSS 2003 study, content and curriculum standards developed by policymakers may not align with the content taught in classrooms (i.e., the intended curriculum does not match the implemented curriculum). Content analyses of curriculum are a necessary part of evaluation (Confrey & Stohl, 2004) and have been included in international studies of student achievement (Ferrini-Mundy & Schmidt, 2005; Mullis et al., 2004). The TIMSS video studies examined curriculum and teaching in action (Hiebert et al., 2003). Other studies report on the cognitive complexity of the mathematics lessons as enacted by teachers (Henningsen & Stein, 1997; Stein et al., 1996; Weiss, Pasley, Smith, Banilower, & Heck, 2003). Within such studies, there is an analysis of mathematics processes (e.g., communication, reasoning) included as essential parts of the curriculum.

Standards for mathematics learning are an example of a descriptive product or representation. The standards document can be a comprehensive description of mathematics students are supposed to learn across grade levels. While a standards document can be much more than a descriptive list, at its most basic level it lays out the expectations for mathematics and the scope and sequence for instruction. The second type of curricular artifact includes instructional materials. The *Standards-based curriculum materials* supported with NSF funding in the 1990s are one example of materials that are intended to reform mathematics teaching in line with mathematics standards. At the time of funding, the standards were the 1989 National Council for Teachers of Mathematics (NCTM) standards. However, the materials are still relevant to the 2000 NCTM standards. More evaluation of the *Standards-based curricula* has been done than for other commercially generated curricula as reported by the National Research Council's report, *On Evaluating Curricular Effectiveness* (Confrey & Stohl, 2004). According to the report, 143 of 192 studies classified as comparative analysis, case, content analysis or synthesis studies were connected to the *Standards-based curricula* developed under NSF



funding in the 1990s. Although some of them have undergone extensive evaluation, more research is still required to understand their impacts on student learning (Confrey & Stohl, 2004) particularly as the materials are revised over time. Some MSP grants include facilitating the adoption and professional development related both to *Standards*-based curricula and other curricula connect to the MSP project. In addition, some MSP grants are collecting student achievement and enrollment data related to such curricula as well as other curriculum they have adapted or developed for their purposes.

### Challenging Curricular Goals and Purposes

Challenging curriculum is an example of a qualitative characteristic measured not only by the materials themselves but also within their system of use. To address the goals for a curriculum, an operational definition of “challenging,” or any other qualitative characteristic, should be established. A difficulty in defining “challenging curriculum” is determining what to measure about the impact of the curriculum on the educational system. The goal may include some identified problem within the classroom system (e.g., underperformance of certain student subgroups, changing mathematics standards, changing teaching practice). The identified problem leads to goals that are used to measure how well the problem has been addressed. The identified problem or goal also leads to relevant theory, knowledge, mathematical content, or pedagogical methods that are then represented in the curricular artifacts. The qualitative characteristic (e.g., challenging) impacts each part of the design process (from goal identification to the design of artifacts).

A prevalent curricular goal is to increase student achievement. Measures of student achievement can provide some indication into a local interpretation of challenging. For instance, course enrollment figures (e.g., enrollment in Algebra I, Algebra II, or AP Calculus) and state-level assessments are used as measures. In most cases, the curriculum design project has identified some problem in student achievement to remedy via initiatives incorporating challenging curriculum. Gaps are identified and measured by test scores, course enrollment, and college enrollment. Bringing underperforming students up to a challenging level is both a measure of the curriculum’s effectiveness and a motivation for making curricular change.

However, does the fact that more students are enrolled in a course identified as challenging mean that students are receiving challenging curriculum? Or are they being taught the content that should be required of every student? Is the course content actually challenging or has a previously challenging course been diluted for the purposes of the new population even though the name is the same? As a research question, the notion of a challenging curriculum is then defined by the local student conditions and characteristics. For example, algebra may be challenging for some but not for all. But, all students need to learn algebra so a



“challenging curriculum” would include the expectation that all students enroll in an algebra course.

### Curriculum System Models

Because there are multiple components behind any curriculum design, development, or implementation initiative, an organizer is needed for analysis and comparison across projects. The other challenge for a program evaluation like the MSP-PE is organizing diverse, locally based efforts to form an image both about what is being accomplished in the program as a whole and to organize individual contributions by grants and projects within grants. For example, a curricular initiative designing new curriculum is at a different phase than a project that is implementing an existing set of materials. A project designing new materials will have different artifacts and needs than a project using existing materials. Alternatively, projects with the same motivation (e.g., equity based) may proceed with different artifacts (e.g., learning units, software, curriculum frameworks) and employ different conceptual systems. The organizing structure proposed for mathematics curriculum analysis can be used to track development and to compare curricula along multiple dimensions. In addition, the structure makes explicit the fundamental assumptions and theories behind curricular design. A curriculum model has three components: a conceptual system, a representation system, and purpose and goals (see Table 1). The conceptual system is enacted or embodied in representational systems with a purpose. The purpose of the system model is to organize the parts of a curricular initiative and to emphasize the interactions between aspects of the evolving project.

Other authors have examined mathematical models including conceptual systems for mathematical ideas or systems (Lesh & Carmona, 2003; Lesh, Cramer, Doerr, Post, & Zawojewski, 2003; Lesh, Doerr, Carmona, & Hjalmarson, 2003). For example, students incorporate multiple mathematical ideas to design a measure that represents “level of roughness” in the context. The measures designed are context dependent (e.g., measuring the roughness of a road is different than metal roughness at an atomic level). The methods for measurement also have a purpose. For instance, the space between peaks, the height of the peaks, or the area of the peaks could all be either important or irrelevant depending on the purposes for the measurement. The model for measuring roughness is then a conceptual system incorporating mathematical, scientific and other systems of knowledge with a representation and a purpose. In a parallel situation, a curriculum has representations (e.g., textbooks, software, teaching practices, assessments), underlying conceptual systems, and purpose. For example, the *Connected Mathematics* series includes materials that represent theory about how students learn. The purpose is to “help students develop understanding of important concepts, skills, procedures, and ways of thinking and reasoning in number, geometry, measurement, algebra, probability

TABLE 1  
Components of a Curriculum Model

Component	Definition	Examples
Conceptual system	The system, theoretical framework, perspective, or way of thinking about curriculum, teaching, learning, or other aspects of the learning system	<ul style="list-style-type: none"><li>■ Constructivism</li><li>● Social constructivism</li><li>● Historical mathematical development</li><li>● Project developed</li></ul>
Representation system	The external artifacts that represent the conceptual system	<ul style="list-style-type: none"><li>● Texts, materials</li><li>■ Standards and frameworks</li></ul>
Purpose & goals	The reason or justification for using the particular representation system or conceptual system for the curriculum	<ul style="list-style-type: none"><li>■ Reduce achievement gaps</li><li>■ Introduce new content</li><li>■ Increase enrollment in advanced coursework</li></ul>
Pedagogical framework	The pedagogical strategies employed within the system	<ul style="list-style-type: none"><li>■ Collaborative learning</li><li>● Direct instruction</li><li>■ Problem-based learning</li></ul>
Content	The mathematics content, skills, and topics incorporated in the system	<ul style="list-style-type: none"><li>■ Algebraic thinking</li><li>● Geometry</li><li>■ Data analysis</li></ul>

and statistics” (<http://www.math.msu.edu/cmp/Overview/Glance.htm>). The accompanying materials are one representation of the purpose. The *Standards*-based curricula have different definitions and focus for their development as well as different interpretations for how to engage students in meaningful mathematics (Putnam, 2003). Other types of curricula may emphasize direct instruction, repetition, or incremental learning according to their perspectives on teaching and learning mathematics.

Modeling language is advantageous for discussing the design aspects of curriculum. The models and modeling language incorporates process and product as well as the representative aspects of a system. The model is the product of the design, in this case, a curriculum. The modeling process is the design and implementation of the curriculum. Both processes are captured in the definition of a curriculum model as the development of models (e.g., the representation) may change over time. The components that change can be tracked using a curriculum model as an analytic tool. As with any organizer, no model encompasses every aspect of the related systems. For example, a coordinate plane represents some characteristics of distance but not others. Conveying motion on a coordinate plane requires another representation system, vectors. The vectors highlight some characteristics of motion in the plane—rates of change and the graph of a function—whereas neglecting others or leaving them open to interpretation by the observer. The vectors are useful for some purposes and not for others. The vectors

do require a mathematical conceptual system and have a purpose. Similarly, for a curriculum model, some aspects of mathematics are highlighted whereas others are neglected. The designers have some purpose and goals behind selection of topics, activities, pedagogical strategies, presentation, format and structures.

Investigating a curriculum system is then a question of examining the curricular model rather than investigating one component of a larger system (e.g., materials in isolation from the conceptual system and purposes). Different models can then be compared and contrasted on the basis of their relation to other models to provide a coherent picture of curriculum models as a whole. Returning to the process of defining challenging curriculum, “challenging” is a qualitative characteristic that is measured in light of other aspects of the model. A curriculum may be challenging by some standards of measurement and not by others, depending on the initial motivation for the curriculum design and the measurement. To understand a curriculum model, I go beyond “challenging” to considering when, where, and for whom a curriculum may be challenging. This is broader than saying a curriculum is or is not challenging and emphasizes the conditions under which the curriculum is challenging (similar to the experienced curriculum as described by Clements, 2002, and others).

## TYPES OF CURRICULUM SYSTEM MODELS

Within the curricular innovations, there are a wide variety of curricular approaches incorporating existing materials and the development of new materials. Based on examination of the MSP program and the literature base, I have identified multiple types of definitions that inform the design, development, and implementation of curriculum and varying perspectives on the notion of “challenging content and curriculum.” There are three perspectives at this point: learner centered, content based, and pedagogically based. The definition drives and motivates the activities of the design project. Although I have identified three perspectives, this is not to say that a particular initiative is limited to one perspective or another; rather, there are multiple drivers behind a curriculum innovation. At the research and investigation level, to compare initiatives, the curriculum system model serves to identify the object of interest for the investigation and the nature of the innovation.

To distinguish between the models, there are a few questions to ask. First, what is the starting point for the curricular development? Is it a particular mathematics topic or concept? Is it a method for teaching mathematics? Second, for each aspect of the curriculum model, what is highlighted? For instance, is a feminist perspective driving the design of gender-neutral activities for mathematics? Is the need for new mathematical content (e.g., data analysis in earlier grades) driving curricular design? What are the fundamental assumptions? Although I distinguish between the models, they are not meant to be mutually exclusive. A project can

seek to simultaneously increase women's interest (i.e., have equity-based models) and introduce new engineering content into a mathematics class (e.g., a content-based model). However, the model categories help to distinguish motivations and assumptions behind curriculum that impact comparing and contrasting them along qualitative dimensions, such as "challenging" or "effective." Also, other qualitatively different models may exist.

## Content-Based Perspective

The first perspective to approach is content based. Content-based initiatives and development place the central focus of the design process on a specific topic area in mathematics. The curriculum is not intended to be comprehensive in the sense that it covers all the mathematics a student should know. Rather, the intent of the curriculum development is to focus on specific content in a new way. The development could include investigating how students learn the content, engaging students in new methods or technology, or employing different presentations of the content. At the center of the design process is the content area (often reflected in the title of the project). For example, Geometer's Sketchpad<sup>®1</sup> (Key Curriculum Press, Emeryville, CA) and the accompanying materials have the goal of introducing and developing geometry content in a dynamic fashion. The curriculum and software development has a particular content area or topic as the focus of the design. That content focus brings a different perspective than the development of the Fathom<sup>®</sup> software (Key Curriculum Press, Emeryville, CA) that focuses on statistical data analysis. The software in each case is a means for understanding a particular content area or topic. As another example, *Number Power* (an elementary *Standards*-based curriculum) focuses on the development of students' understanding of numbers (Battistich et al., 2003). The nature of the content is a driver behind the curriculum design. The curricular initiative in the case of Geometer's Sketchpad, Fathom, and calculus reform is centered on particular content. Hence, a primary question to ask about a curriculum initiative is what content is being emphasized (or not). Following the descriptive question, what about the nature of the content is motivating the design of the curricula in particular directions? Understanding what content is emphasized or not aids the evaluation. In addition, understanding that a project is content-focused changes the type of comparison that may be done with curricular initiatives that may have other emphases.

From a modeling framework, the content-based perspective implies purposes that are driven by the content area (e.g., investigate whole number concepts). The representations will vary by the content area. In the case of Geometer's Sketchpad, the dynamic drawing tools for constructing geometric figures are central to the curriculum. For *Number Power*, representations present views of number

---

<sup>1</sup>For more information, see <http://www.keypress.com/sketchpad/>.



and operations. The representations are tied to the conceptual understanding the designers are trying to develop in and for students. The theoretical perspectives, theories, and knowledge that are incorporated for the curriculum design rely on knowledge about students' development of conceptual understanding in the topic area. However, a motivation for a curriculum design project could be to learn about the development of a mathematics concept when little to no research base may be available. For example, Rasmussen and colleagues have designed materials for teaching undergraduate differential equations that have resulted in research-based knowledge about how students understand and learn about differential equations (Rasmussen, 2001; Rasmussen & Stephen, 2002). The design of the materials operates in concert with the development of knowledge about learning the content.

There are possible subgoals within a content-based perspective. The first possible subgoal is the introduction of new content that was not previously part of the standards for mathematics. The content is "challenging" in the sense that it is emerging, cutting-edge content that may even be new for experts in the field. For example, recent engineering-based projects are using nanotechnology content as part of the modules and units designed for classroom use. Nanotechnology is a new and emerging field within engineering, science and technology. In this sense, the content is identified as challenging because of its emerging and complex nature. Such curricular initiatives begin at earlier stages of the curricular design process because the nature of the content may not be clear. Understanding how students learn such content may be part of the curriculum design initiative (Burkhardt et al., 1990). For Clements's (2007) curriculum research framework, the investigation lies in the early stages where researchers are developing theory about learning and teaching. Because the content is new, there is unlikely to be research about students' learning of the content to drive curricular design. The designers are then learning about students' understanding as they develop the curriculum.

A different type of subgoal focuses development not on new content but looking at existing content from a new perspective. For instance, the Fathom software gives a presentation of ideas such as mean and standard deviation in a dynamic way that was not possible with paper-based calculations. Although the content is not new, the students may learn about the content in a fundamentally different way. This creates a contrast to initiatives that are developing materials for content already within a typical mathematics curriculum (e.g., ratios, fractions, operations). Other examples of new perspectives on content include research on computer algebra systems or graphing calculator applications. All of these are curricular initiatives that approach existing concepts from new perspectives. Again, students' learning and understanding of the content may be different leading to questions about how the innovation has changed student learning and how that learning should be assessed in the new context.

Overlap with a pedagogical or equity-based perspective is possible here as new approaches to existing content may be necessary to address inequities between

subgroups. In a pedagogically based perspective, the focus shifts from content to teaching. Although teaching and teaching methods are relevant to discussions of content, the pedagogically based initiative may not focus on a particular content area.

### Pedagogically-Based Perspective

Distinct from the content of the curriculum is how the content is taught, as well as the principles behind the teaching. A pedagogically based perspective on curriculum emphasizes the teaching methods or strategies that are used to introduce the content. For example, there is a difference in challenge between being shown an algorithm for adding fractions and developing an algorithm for fractions. The pedagogy for engaging in each type of activity is different. For instance, the *Standards*-based curricula emphasize particular methods for teaching (e.g., collaborative learning, manipulatives) and presentation of content by particular types of classroom activity intended to develop conceptual understanding. Other curricula may emphasize the division of concepts into subtopics explicitly described to students. There are particular pedagogical theories that inform the design of a curriculum that develops particular content but are not necessarily motivated by the type of content. Assumptions about how mathematics should be taught are embedded in the curriculum design and implementation that may be independent from the mathematics topics (e.g., encouraging mathematical reasoning and discourse).

In a pedagogically based curriculum model, even though the content may not be new, it is presented in a different way from a pedagogical perspective. Although teachers are a critical component of any curricular initiative, the teachers are particularly important. Their beliefs, understandings, and interpretations of students, mathematics, learning, and teaching play a significant role in the ultimate implementation of the curriculum. To return to MacNab's (2000) three-stage curriculum framework, the match between the ideal and the implemented curriculum depends heavily on the teacher.

A challenging curriculum from a pedagogically based perspective then uses pedagogical strategies that challenge students to develop knowledge, to work on fundamental mathematics, to engage in high-level mathematical work. The implication of a challenging, pedagogically based curriculum is pedagogy that engages students in challenging mathematical work. For a teacher, this may mean their own knowledge of mathematics needs to be well developed to engage in challenging pedagogy. The mathematics teachers need is distinct from the mathematics needed for other professions and is focused on students' ways of thinking and learning (Hill, Rowan, & Ball, 2005). The teachers' mathematical knowledge impacts pedagogy in terms of the types of activities that are then possible (Remillard, 2000). For example, engaging students in activities that require them to describe proofs requires the teacher to also be able to assess and evaluate those proofs. This

goes beyond analyzing any one mathematical argument to analyzing multiple arguments as well as organizing mathematical discourse effectively.

The pedagogically based perspective focuses on the teaching methods used in association with curriculum. The pedagogy may intend to address the learning needs of some subgroups of the student population. Following MacNab's (2000) framework from intended to experienced, I shift the discussion at this point to learner-centered perspectives that focus their work at learner-related goals with supporting work at the teacher level. The type of gaps and the subgroups identified may vary by curriculum initiative, but the intent to address inequity implies a different type of model than a strictly pedagogical or content-based initiative.

### Learner-Centered Perspective

Learner-centered models include representations, concepts, and purposes related to learners as the focus of project activity. For instance, "increasing the college enrollment and retention rates" is a learner-centered activity. This is not to say that learners are not a component of pedagogical systems or content-based systems rather that learner-centered models place learners at the center of project activity and work with teachers and content may be in place to support the learner-centered objectives. A learner-focused perspective may approach curriculum systems as the solution to reducing inequity (e.g., in achievement, in retention, in recruitment) between subgroups of the student population. From a position of equity, the designers identify the subgroups to be impacted (e.g., inequity based on language differences requires a different approach than inequity based on gender), the content to be addressed, and the pedagogical approaches for the curriculum. For instance, encouraging the retention of women in engineering was the focus of a curriculum-design project at Purdue University. Curricular modules were developed with the intent of reducing gaps in enrollment between women and other populations. The curriculum development impacted the whole population and the education system as a whole (Diefes-Dux, Follman, Zawojewski, Capobianco, & Hjalmarson, 2004). With a similar goal, Carnegie Mellon University redesigned its computer science program based on research about women's perceptions of computer science and the characteristics of computer science majors (Margolis & Fisher, 2002). Although both projects ultimately affected all students, the motivation for the curricular reform was founded on an equity goal. Within MSP grants, examples exist where projects are utilizing after-school activities, mentoring programs, and other initiatives across K-16 to encourage students (particularly from underrepresented populations in science, technology, engineering, and mathematics professions) to complete higher education. Content and pedagogy are then considered from the perspective of reducing gaps based on some measure of equity.

Impacting learner-centered models are calls from the No Child Left Behind Act (NCLB) for reduction in achievement gaps between subgroups. There is



debate about the measures used to measure and to quantify whether gaps between subgroups exist or have been eliminated. Kim and Sunderman (2005) presented alternative measures of schools' progress toward reducing gaps and a description of the differential impacts of NCLB on high-poverty schools, based on the use of the mean as an absolute measure of yearly progress. NCLB aside, other equity-based measures of challenge include Advanced Placement course enrollment and enrollment in gateway courses such as algebra. The challenge in a learner-centered model (especially related to equity concerns) perspective is to determine what reforms need to occur to reduce gaps in achievement, as well as how those gaps will be measured over time to determine progress toward the goal. As an example, measuring high school dropout rates is one indicator of student performance. However, because of student mobility and opportunities to take the GED, it is not always clear how to count which students are dropouts and which have pursued alternatives to a high school diploma. To refer back to a curriculum model, gaps in achievement are both motivation as well as measure of success as they are reduced over time.

Learner-centered perspectives on challenging curriculum can go beyond the examination of gaps in achievement at the district or school levels to the design of curriculum within local contexts. Gutstein's (2003) study of his urban, Latino students' interaction with *Standards*-based curriculum materials and the design of projects specifically for his student population brings up a version of challenging curriculum that is both challenging in the sense of presenting complex content but also meaningful to the students. His students were challenged to investigate questions about poverty and socioeconomic status that were personally meaningful. The *Standards*-based curriculum alone may not have been challenging to the population without the use of supplementary investigations. In this sense of challenging, the students' understanding of their world was pushed and they were expected to develop arguments and use mathematics to achieve greater understanding of meaningful questions. In this type of learner-centered model, the locally available curriculum encourages the investigation of personally meaningful questions while developing and applying significant mathematics. The curriculum in this context is then an integration of the mass-produced materials and locally developed materials. How the materials and activities are integrated to develop a locally meaningful, challenging curriculum requires study of teachers' design of curriculum for their classroom and students' experiences as learners. In MacNab's (2000) framework, the teacher will have an ideal curriculum based on understanding of the local population. The teacher documents the implementation and learns what the students' experiences were to inform future instruction.

## CONCLUSION

A theme for curricula is their nature as a system of interacting components. The model is used as an analytic structure to categorize, classify, and distinguish types



of curriculum systems by their purposes, representations, and underlying conceptual systems or theories. The three models presented (pedagogically based, content based, and equity based) are three windows or entry points into curriculum. There are points of intersection between them (e.g., what are pedagogical models that impact equity-based curriculum design?). All three focus on the purposes, goals, and representations of curricula while approaching curricula from different points. Beyond these three types, other models and perspectives may exist that could be described using a curriculum model as a perspective. In addition, there may be subclasses such as equity-focused learner-centered models (e.g., curricular initiatives focused on the increasing representation in science, technology, engineering, and mathematics fields of underrepresented groups).

As an analytical frame, curriculum models unify aspects of curricula that create distinctions between projects. These include the purposes, representations, and conceptual systems behind the design and implementation of the curriculum. To return to definitions of challenging curriculum, the definition and measurement of challenging depends on the particular curriculum models employed by the project. Measures of challenge and meeting goals related to the implementation of challenging curriculum then depend on the particular curriculum model. Rather than presenting one measure of challenging, the use of curriculum models provides a means for examining multiple measures and comparing curricula based on similar models (e.g., equity-based curriculum efforts can be compared with other equity-based curriculum efforts).

The use of the language of modeling affords focus on both the product and the design process by examining the representations and purposes for curriculum closely and highlighting the features that go beyond the materials themselves to the contexts and conditions of their use. The representational nature is not meant to be static. Rather, the usefulness of the representation changes over time and with experience. In short, what is challenging today may not be challenging tomorrow. What is challenging curriculum in one context may not be challenging in another context. The model structure allows for classification across types, but accounts for differences at the local implementation level. The language of modeling also fits with curriculum frameworks that focus on the intended, the implemented or enacted, and the experienced curriculum that look at different stages in the curriculum system.

## ACKNOWLEDGMENT

This study is one in a series of substudies for the Math and Science Partnership Program Evaluation (MSP-PE) conducted for the National Science Foundation's MSP Program. The MSP-PE is conducted under Contract No. EHR-0456995. Since 2007, Bernice Anderson, Ed.D., Senior Advisor for Evaluation, Directorate for Education and Human Resources, has served as the National Science

Foundation Program Officer. The author, from George Mason University, is Margaret A. Hjalmarson, Ph.D.

The MSP-PE is led by COSMOS Corporation in current partnership with George Mason University (GMU) and Brown University. Robert K. Yin (COSMOS) serves as Principal Investigator and Jennifer Scherer (COSMOS) serves as one of three Co-Principal Investigators. Additional Co-Principal Investigators and their collaborating institutions (including discipline departments and math centers) are Patricia Moyer-Packenham (USU, formerly GMU) and Kenneth Wong (Brown).

## REFERENCES

- Battistich, V., Alldredge, S., & Tsuchida, I. (2003). Number power: An elementary school program to enhance students' mathematical and social development. In S. L. Senk & D. R. Thompson (Eds.), *Standards-based school mathematics curricula: What are they? What do students learn?* (pp. 133–160). Mahwah, NJ: Erlbaum.
- Burkhardt, H., Fraser, R., & Ridgway, J. (1990). The dynamics of curriculum change. In I. Wirszup & R. Streit (Eds.), *Developments in school mathematics around the world* (Vol. 2, pp. 3–30). Reston, VA: National Council of Teachers of Mathematics.
- Clements, D. H. (2002). Linking research and curriculum development. In L. D. English (Ed.), *Handbook of international research in mathematics education* (pp. 599–630). Mahwah, NJ: Erlbaum.
- Clements, D. H. (2007). Curriculum research: Toward a framework for "research-based curricula." *Journal for Research in Mathematics Education*, 35(1), 35–70.
- Confrey, J., & Stohl, V. (Eds.). (2004). *On evaluating curricular effectiveness: Judging the quality of K-12 mathematics evaluations*. Washington, DC: National Academies Press.
- Diefes-Dux, H., Follman, D., Zawojewski, J., Capobianco, B., & Hjalmarson, M. (2004). *Model eliciting activities: An in-class approach to improving persistence and retention of women in engineering*. Paper presented at the annual conference for the American Society of Engineering Education, Salt Lake City, Utah.
- Ferrini-Mundy, J., & Schmidt, W. H. (2005). International comparative studies in mathematics education: Opportunities for collaboration and challenges for researchers. *Journal for Research in Mathematics Education*, 36(3), 164–175.
- Gutstein, E. (2003). Teaching and learning mathematics for social justice in an urban, latino school. *Journal for Research in Mathematics Education*, 34(1), 37–73.
- Harris, K., Marcus, R., McLaren, K., & Fey, J. T. (2001). Curriculum materials supporting problem-based teaching. *School Science and Mathematics*, 101(6), 310–318.
- Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education*, 28(5), 524–549.
- Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., et al. (2003). *Teaching Mathematics in Seven Countries: Results from the TIMSS 1999 Video Study* (No. NCES 2003-013 Revised). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2), 371–406.
- Hjalmarson, M. (2005). Purposes for mathematics curriculum: Pre-service teachers' perspectives. In G. M. Lloyd, M. Wilson, J. L. M. Wilkins & S. L. Behm (Eds.), *Proceedings of the 27th Annual*

- Meeting of the International Group for the Psychology of Mathematics Education—North American Chapter [CD-ROM]*. Eugene, OR: All Academic.
- Kim, J. S., & Sunderman, G. L. (2005). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational Researcher*, 34(8), 3–13.
- Lesh, R., & Carmona, G. (2003). Piagetian conceptual systems and models for mathematizing everyday experiences. In R. Lesh & H. M. Doerr (Eds.), *Beyond constructivism: A models & modeling perspective on mathematics problem solving, learning & teaching* (pp. 71–96). Mahwah, NJ: Erlbaum.
- Lesh, R., Cramer, K., Doerr, H. M., Post, T., & Zawojewski, J. S. (2003). Model-development sequences. In R. Lesh & H. M. Doerr (Eds.), *Beyond constructivism: A models & modeling perspective on mathematics problem solving, learning & teaching* (pp. 35–58). Mahwah, NJ: Erlbaum.
- Lesh, R., Doerr, H. M., Carmona, G., & Hjalmarson, M. (2003). Beyond constructivism. *Mathematical Thinking and Learning*, 5(2,3), 211–234.
- Lloyd, G. M. (1999). Two teachers' conceptions of a reform-oriented curriculum: Implications for mathematics teacher development. *Journal of Mathematics Teacher Education*, 2, 227–252.
- Lloyd, G. M. (2002). Mathematics teachers' beliefs and experiences with innovative curriculum materials: The role of curriculum in teacher development. In G. C. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: A hidden variable in mathematics education* (pp. 149–160). Dordrecht, The Netherlands: Kluwer Academic.
- MacNab, D. (2000). Raising standards in mathematics education: values, vision, and TIMSS. *Educational Studies in Mathematics*, 42(1), 61–80.
- Margolis, J., & Fisher, A. (2002). *Unlocking the clubhouse: Women in computing*. Cambridge, MA: MIT Press.
- McCaffrey, D. F., Hamilton, L. S., Stecher, B. M., Klein, S. P., Bugliari, D., & Robyn, A. (2001). Interactions among instructional practices, curriculum, and student achievement: The case of standards-based high school mathematics. *Journal for Research in Mathematics Education*, 32(5), 493–517.
- Middleton, J. A. (1999). Curricular influences on the motivational beliefs and practice of two middle-school mathematics teachers: A follow-up study. *Journal for Research in Mathematics Education*, 30(3), 349–358.
- Mokros, J. (2003). Learning to reason numerically: The impact of investigations. In S. L. Senk & D. R. Thompson (Eds.), *Standards-based school mathematics curricula: What are they? What do students learn?* (pp. 109–132). Mahwah, NJ: Erlbaum.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.
- National Council for Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.
- National Council for Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. Reston, VA: Author.
- Pligge, M. A., Kent, L. B., & Spence, M. S. (2000, April). *Examining teacher change within the context of mathematics curriculum reform: Views from middle school teachers*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31(7), 3–14.
- Putnam, R. T. (2003). Commentary on four elementary mathematics curricula. In S. L. Senk & D. R. Thompson (Eds.), *Standards-based school mathematics curricula: What are they? What do students learn?* (pp. 161–180). Mahwah, NJ: Erlbaum.
- Rasmussen, C. L. (2001). New directions in differential equations: A framework for interpreting students' understandings and difficulties. *Journal of Mathematical Behavior*, 20, 55–87.

- Rasmussen, C. L., & Stephen, M. (2002). Classroom mathematical practices in differential equations. *Journal of Mathematical Behavior*, 21, 459–490.
- Remillard, J. T. (2000). Can curriculum materials support teachers' learning? Two fourth-grade teachers' use of a new mathematics text. *The Elementary School Journal*, 100(4), 331–350.
- Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H., Wiley, D. E., Cogan, L. S., et al. (2001). *Why schools matter: A cross-national comparison of schools and learning*. San Francisco, CA: Jossey-Bass.
- Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in classrooms. *American Educational Research Journal*, 33(2), 455–488.
- Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. J. (2003). *Looking inside the classroom: A study of K-12 mathematics and science education in the United States*. Chapel Hill, NC: Horizon Research.



# A Review of Instruments to Evaluate Partnerships in Math and Science Education

Jennifer Scherer  
*COSMOS Corporation*

The importance of the role of partnerships in research is evidenced by the federal government's, private sector's, and nonprofit organizations' continued interest in and approach to funding research through these vehicles. In education research involving interorganizational partnerships, partnerships are needed to create coordination and alignment across institutions of higher education as well as within K-12 systems. Successful partnership building requires significant resources in terms of human effort and dollars spent. It is therefore critical that partnerships evaluate themselves and their activities. This article provides a description of and reviews instruments that measure different aspects of partnerships and further suggests that instead of using any instrument in toto, that it be modified for evaluation of specific traits of a partnership and validated in the local context. The article further provides an illustrative example of educational evaluation from the National Science Foundation's Math and Science Partnership Program, which calls for interinstitutional partnerships among institutions of higher education, local education agencies, state education agencies, and other for-profit and nonprofit entities.

Interorganizational partnerships are an important part of research endeavors that anticipate having relevance and real-world applicability. Partnerships can create lasting linkages and accountability among those involved and contribute to sustaining the research outcomes. Approaching research individually, partners are less likely to accomplish what would otherwise be possible as a collective. In other words, the partnership as a unified entity is able to generate more than the sum of its partners.

Overall, however, there is still insufficient empirical evidence on how partnerships work and if they result in their intended outcomes (Marra, 2004), but there is

evidence that partnerships create added value (Barnes, Carpenter, & Bailey, 2000). In addition, there has been relatively little attention given to how to effectively evaluate partnerships in the general sense (Schulz, Israel, & Lantz, 2003). Successful partnership building requires a significant amount of time, money, and human effort—all of which may be considered precious resources to partnerships with limited budgets. Partnerships can use information obtained from a self-evaluation to improve their performance and overall operations. Relatively few instruments exist that measure the totality of a partnership, but many instruments have measures for specific aspects, such as measuring trusts or mutuality. Given that partnerships vary dramatically in their origins, start-up and organizational activities, research efforts, and intended outcomes, this article proposes that partnerships adapt existing instruments to meet their evaluation needs rather than developing a new instrument or foregoing evaluation entirely. Furthermore, it will be necessary to gather validity evidence relevant to the intended inferences and uses for the local context.

For K-16 public education, the main hypothesis is that partnerships are needed to coordinate and align the actions and policies leading to improved student achievement—starting with widespread agreement over the goals for student learning, based on rigorous content and performance standards (e.g., Raizen, McLeod, & Howe, 1997). The need for coordination and alignment reflects an essential aspect of K-16 student achievement; such performance ultimately results from a complexity of institutions (not just the formal K-16 system, and most certainly not just what takes place in a K-16 classroom):

- Family and community institutions that heavily define students' learning experiences outside of the school (e.g., learning at home; after-school programs)
- Business and job markets that create opportunities and expectations for students during and after their academic careers (Vinten, 1996)
- College admissions criteria that serve as a highly motivating force for pre-collegiate schooling (e.g., Callan, 1998; Langenberg, Marx, & Shapiro, 1999)
- Teacher preparation and professional development offerings by institutions of higher education (IHEs) that affect the quality and quantity of a student's teachers
- A whole host of policies implemented by state departments of education, regarding student promotion, course requirements, assessments, and curriculum, as well as teacher certification rules (e.g., Teitel, 1993)

Partnerships are needed to create coordination and alignment across these institutions, as well as within K-16 systems that traditionally have been "loosely-coupled" (Weick, 1976). Partnerships also can provide continuity of focus, align curricula and assessments, create desired normative climates, and instill

accountability (Elmore, 2000). For example, the Annenberg Foundation's "Challenge" gifts, which began in 1993, have helped build strong coalitions among businesses, foundations, universities, and grassroots community groups to muster greater public will and support for public school reform (The Annenberg Foundation, 2002).

At the same time, previous research suggests that collaborations between IHEs and local education agencies (LEAs), far from taking place within a congenial framework, may even evoke the clashing of two cultures (Committee on SMTP, 2001; Conference Board of the Mathematical Sciences, 2001; Goodlad, 1993; Goodlad & Sirotnik, 1988). Some of the participating IHEs might even have grappled with the historic role of schools of education (Clifford & Guthrie, 1988; Tierney, 2001; Timpane & White, 1998) and the evolving role of "professional development schools" (Clark, 1999; Committee on SMTP, 2001; Holmes Group, 1990; Rice, 2002). Given the nuances of partnering among educational institutions and agencies, the importance of evaluation becomes evident.

This article provides a description of and reviews instruments from a wide variety of fields that measure different aspects of partnerships. Given that a significant number of partnership evaluation instruments specific to education do not exist, the article posits that instruments may be adapted from fields other than education to evaluate education partnerships. This can be accomplished by utilizing components of the instruments relevant to a particular partnership. The article further provides an illustrative example of educational evaluation from the National Science Foundation's (NSF's) Math and Science Partnership (MSP) Program, which calls for interinstitutional partnerships among institutions of higher education, local education agencies, state education agencies, and other for-profit and nonprofit entities. In reviewing evaluation and assessment partnerships amongst education institutions and agencies, NSF's MSP Program provides an illustrative example of how these partnerships are addressing questions of self-evaluation and evaluation of partnership outcomes.

## EDUCATION PARTNERSHIPS IN THE MSP PROGRAM

The MSP Program at NSF promotes the development, implementation, and sustainability of exemplary partnerships to produce high-quality math and science education at all K-12 levels. The MSP Program anticipates that the partnerships will be instrumental in improving student achievement, as well as reducing achievement gaps among diverse student populations differentiated by race/ethnicity, socioeconomic status, gender, or disability, a strategy advocated by Haycock, Hart, and Irvine (1992). The importance of being partnership driven with science, technology, engineering, and math faculty engagement is apparent not only from the name of the program but also in NSF's decision to include it as one of the five

“key features.”<sup>1</sup> Given the fundamental importance of these partnerships, how the partnerships are evaluating themselves as a functioning entity, becomes a critical question.

The complexity of the MSP Program derives both from the nature of the individual grants and their collectivity. Individually, each of the grants is being conducted by a partnership and not a single entity, with a core set of partners deeply engaged in the effort at both institutional and individual levels—sharing goals, responsibilities, and accountability for the grant.<sup>2</sup>

A required partnership in the MSP Program is between an IHE or eligible nonprofit organization (or consortium of such institutions or organizations) and one or more LEAs that may also include a state educational agency or one or more businesses.<sup>3,4</sup> This type of partnership arrangement is vertical in nature in that LEAs are partnering with entities (e.g., IHEs) at later points along the education continuum.<sup>5</sup> This verticality may enable the LEAs to maximize their educational potential and establish student pathways (Howard Community College, 1999). The MSP Program also distinguishes between core and noncore partners. Core partners share responsibility and accountability for the MSP grant. All core partner organizations are required to provide evidence of their commitment to undergo the coordinated institutional change necessary to sustain the partnership effort beyond the funding period. A noncore or supporting partner is *not* required to commit to the institutional change necessary to sustain grant activities beyond the funding period but is an important stakeholder in K-12 math and science education.

## Partnership Assessment Efforts in the MSP Program

As of the grant period 2002–2003, MSP Program partnerships were in an initial evaluation-planning and implementation phase. As illustrative examples of these initial activities, one partnership worked with a foundation to develop a partnership evaluation instrument. Another partnership developed evaluation questions that included (a) to what extent the partnership is using existing resources and lessons

---

<sup>1</sup>The four remaining key features include (a) teacher quality, quantity, and diversity; (b) challenging courses and curricula; (c) evidence-based design and outcomes; and (d) institutional change and sustainability.

<sup>2</sup>In reviewing a sample of MSP grants, all of the partnerships, except for one, enacted the partnership with the partners originally proposed. In the one partnership that differed from the original set of partners proposed it enacted one district-level partner that was not proposed and did not enact one district-level partner that was proposed.

<sup>3</sup>National Library of Congress, *National Science Foundation Authorization Act of 2002* (Public Law 107-368), U.S. Government Printing Office, Washington, DC.

<sup>4</sup>Many good institutional partnerships are driven by strong interpersonal relationships within the institutional partnerships. The interpersonal relationship may have been the original driver, but there is a real need for interpersonal and institutional connectivity.

<sup>5</sup>There is no intended value in the continuum (e.g., from good to bad or vice versa).



from previous initiatives to their advantage, (b) how efficiently and effectively the partners work together, and (c) to what extent the resources and capacities of the partners is adequate for carrying out goals with quality. Another partnership was developing a partnership evaluation instrument to include indicators designed to examine the process of building a functional and healthy relationship.

One partnership planned to measure the degree to which a true effective partnership was established and to identify the defining attributes of such a partnership. Another planned to evaluate the efficacy of the partnership. A final partnership planned to evaluate cultural changes within the participating institutions, including reward systems, district priorities and policies, IHE priorities and policies, and lines and type of communication and participation. Other partnerships have indicated that they will be performing evaluations of the partnership in subsequent years of the grant.

At the same time, some of the partnerships developed plans to disseminate findings about their partnerships, including providing documentation of what works, and providing information about how to construct such a partnership, to a wide audience of policymakers and university and school leaders. Another partnership read and discussed "Effective School-College Partnerships, A Key to Education Renewal and Instructional Improvement" to increase their understanding of partnerships and to assess their prior interactions against the described criteria to identify strengths and areas for improvement. A final partnership reported that its advisory board would provide comment on the general progress and direction of the partnership.

These ongoing experiences with the MSP Program's partnerships suggest that they, as well as possibly other partnerships, lack familiarity with and access to the needed evaluation instruments. The remainder of this article therefore reviews an array of such instruments. The review is organized according to a generic partnership framework, to help existing and future partnerships to identify the instruments that best match a partnership's evaluation priorities.

## PARTNERSHIP EVALUATION AND INSTRUMENTS

An understanding and organization of existing partnership instruments from a range of different fields may allow partnerships to more readily access evaluation instruments appropriate for their needs. To both understand and organize a sample of the existing instruments a basic framework is needed that exemplifies the typical phases of a partnership. Extant research appears to support and be consistent with a general framework as depicted in Exhibit 1, which highlights the typical phases of an Interorganizational partnership.

The following section reviews assessment methods and instruments and uses the framework as a guideline to examine how different instruments can be used to look at the different phases of a partnership. This review is based on a comprehensive

search and review of the extant literature on partnerships, partnership evaluation, and partnership-based research. The search included a review of documents from several different disciplines, many from health-oriented fields. Many of the education-related partnerships appeared in the area of informal education, or other areas such as education of patients or students from medical or clinical studies (Sackett, Hendricks, & Pope, 2000; Warrick, Wood, Meister, & de Zapien, 1992), or education of participants in a study on a particular topic such as social inclusion (Clegg & McNulty, 2002). This review focuses only on those sources that included partnership instruments or performance indicators; I also contacted the authors of the many works cited to obtain a copy of the instrument for an in-depth examination.

Assessment Methods and Instruments

Some instruments exist for assessing and evaluating different aspects of partnerships and other partnership-related entities such as community coalitions or organizations, and additional instruments may be adapted and modified for evaluation purposes (see Figure 1).

The instruments come from a wide range of disciplines but are arguably adaptable for partnerships along many different dimensions. Systematic evaluation of partnerships will help to determine (a) the basic characteristics of the partnerships, (b) how partners work with each other, (c) the dimensions of the partnering relationship, and (d) immediate effects of the partnership on the partners.

In the long term, it is essential to determine the effect of the partnership on the intended outcomes (such as changes in student achievement). However, it is difficult to determine the extent to which partnerships actually work if the only outcome studied is distal in nature and if no theoretical link has been established between the partnership and its long-term outcomes. In fact, much of the evidence about partnerships' contributions to overall performance is, with the exception of

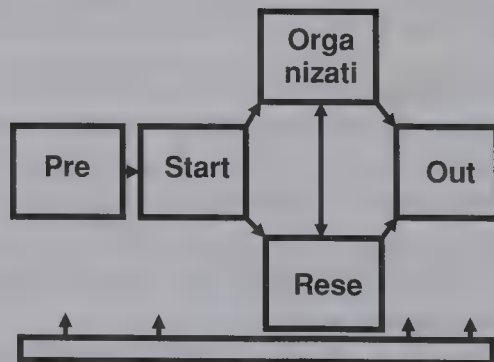


FIGURE 1 A partnership framework.

a few private sector alliances, where increased efficiencies have been documented and quantified anecdotally (Shah & Singh, 2001).

To attribute distal outcomes to the work of the partnership, it is important to have documented the partnership start-up process, identified key elements of the partnering relationship, and assessed the immediate effects of the partnership on major stakeholders: the members of the partnership, the partnership itself, and the targeted community. In the case of an IHE–K-12 partnership, the members are (a) the researchers, faculty, and administrators at the university, as well as the students (who may be termed “service learners”); (b) the K-12 teachers, administrators, and students; (c) other partnering organizations and community advocates; and (d) the members of the targeted community.

The literature does not uniformly support any particular instrument for comprehensive evaluation of the types of partnerships in the MSP Program. However, there are a number of instruments that may be used as is (with local validation) or adapted and validated for evaluation of the various components of the MSP Program partnerships. Selection of instruments should be based on the purposes, goals, and objectives of the partnership and can be adapted to its particular context. Methods used to administer the instruments include in-person interviews with partners, partner surveys, and observations. As shown in Table 1, most of the instruments discussed cover more than one dimension of working with partners and can be categorized as working with partners, partner relationships, and increased capacity.

To assess the working of partnership-based community coalitions, the John S. and James L. Knight Foundation and The Robert Wood Johnson Foundation developed a number of survey instruments: a nine-question expert advisory panel instrument, an 18-question mail survey, a 45-min telephone survey for leaders, a 20-min telephone survey of key informants (e.g., nonleaders), and an in-depth site visit guide (Drug Strategies, 2001). These instruments are comprehensive in nature and vary in topical areas covered and length.

### Pre-Existing Partnership(s) and Start-Up of Partnership

To measure partnerships, Kingsley and O’Neil (2004) developed a three-staged partnership logic model. Stage 1, partnership preconditions, examines the embeddedness of the partnership. Kingsley and O’Neil defined embeddedness as the number and types of relationships that organizations have with one another prior to the development of the partnership. Stage 1 further explores the strategic needs or the types of resources and needs confronting organizations as well as whether there is a congruence or complementarity in these needs. Stage 2, partnering activities, looks at partnership formation (including aspects such as agreements, goals, resource allocation, etc.) and partnership operations or the actual behaviors in which the partners engage. Stage 3, partnership outcomes, examines both process and performance outcomes. Kingsley and O’Neil defined process outcomes as

TABLE 1  
Partnership Dimensions Addressed by Specific Instruments

Instrument	Source	Working With Partners					Partner Relationships			Increased Capacity			
		Community Readiness	Recruiting Partners	Missions and Goals	Original Structure	Action and Plan	Finances		Evaluation Monitoring	Mutuality Trust	Leadership Collaboration	Partner and Partnership Capacity	Community Capacity
							Resources	Technical Assistance					
A collaboration progress checklist	Bergstrom et al. (1995)	X		X					X		X	X	
Campus community partnership principles	Bell-Elkins (2002)		X	X					X		X		
Capacity: a collaborative assessment	Gardner (1995b)				X	X			X		X	X	
Collaboration	Hays et al. (2000)										X		
Collaboration of assessment tool	Borden and Perkins (1999)					X			X		X		
Collaborative factors inventory	Mattessich et al. (2001)	X	X								X		
Collaborative values	Gardner (2000)	X									X		
Community empowerment	Bartle (2003)		X	X	X					X	X	X	X
Community-based organizational self-assessment tool	Bright (1998)		X	X	X	X			X		X		
Effectiveness of community coalitions	Butterfoss et al. (1996b)		X										
Establishment and effectiveness of state minority health networks	van Houten et al. (2000)		X	X	X	X			X				
Institutional learning	Rothwell (2004)											X	
Interview guide for expert advisory panel	Drug Strategies (2001)	X											
Mail survey for coalition member					X	X	X	X					





“[1] “the qualitative and quantitative assessments that measure whether the partnership achieved the goals and duties of operation” and performance outcomes as “improvements . . . in the working environments of the organizations, 2) the transfer of knowledge between organizations, and 3) the increased ability to quickly innovate” (Kingsley & O’Neil, 2004).

Butterfoss, Goodman, Wadersman (1996a) developed a Plan Quality Index. The instrument includes a preimplementation checklist and examines respondents’ assessment of the components of a committee plan: goals and objectives, scope, and community resources.

Gardner (1995b) developed a 29-item Community-Based Self-Assessment Instrument that measures nine dimensions of a community organization’s progress toward responding to policy changes. The nine dimensions are (a) collaboration with other agencies, (b) internal agreement on values and mission, (c) diversity and inclusiveness, (d) organizational priorities, (e) budgets and resources, (f) staff and leadership development, (g) commitment to outcomes and accountability, (h) response to policy changes, (i) future role of the organization.

## Organizational Structure and Ongoing Partnership Operations

Mutuality and trust are two critical elements of successful partnership structure and operations, and therefore should be part of the measurement process. Next are measures that specifically focus on mutuality and trust, leadership, and collaboration. The following describes instruments that measure trust from a management perspective and may be effective in measuring these aspects of the MSP Program partnerships.

### Mutuality and Trust

The U.S. Department of Education’s Institute of Education Sciences holds that a key item in a study analyzing outcome data is that the measures are valid, that is, they accurately measure the true outcomes that the activity was designed to affect (U.S. Department of Education, 2003). Metzler, Higgins, Beeker, Freudenberg, Lantz et al. (2003) searched for validated instruments to measure trust between community and research partners and were unable to find any.

Paine (2003) presented a trust measurement questionnaire intended to answer the following three questions: (a) Have the organization’s programs and activities changed what people know, think, or feel about the organization, and how they act; (b) have the actions of the organization had an impact on how constituents trust the organization; and (c) can the organization document that its communication efforts have increased this trust? The instrument covers mutuality, commitment, satisfaction, communal relationships, and exchange relationships.

Communal relationships are those in which both parties provide benefits to each other; in exchange relationships, one party gives benefits to the other, because the other party has done so in the past or is expected to do so in the future. According to Paine, "communal relationships are essential to developing and enhancing trust in an organization."

Glaeser, Laibson, Scheinkman, and Soutter (2000) combined two experiments and a survey to measure trust and trustworthiness, which they define as two key components of social capital. Gillespie (2003) developed a behavioral trust inventory based in part on existing measures of trustworthiness, disposition to trust, trust in the team, common values, common goals, interdependence, risk in the relationship, relationship effectiveness, overall trust, strength of the relationship, and satisfaction with performance. Lantz, Viruell-Fuentes, Israel, Softley, and Guzman (2001) found in their evaluation that building trust among partners is an important factor for growth and achievement in partnerships.

## Leadership

Hays, Hays, DeVille, and Mulhall (2000) studied the relationship between the structure of substance abuse prevention coalitions and community impact. They measured leadership effectiveness through a six-item instrument assessing members' perceptions of the extent to which the coalition leader directs the group toward collaborative goal achievement. Each item was measured on a 5-point Likert scale. Goodman, Wandersman, Chinman, Imm, & Morrissey (1996) developed another instrument, a key leader survey. Timperley and Robinson (2002) suggested that leadership figures should be aware of the critical importance of partnerships to schools, agencies, and communities. Lantz et al. (2001) found that successful partnerships garnered committed and active leadership from community partners.

## Collaboration and Communication

Gajda (2004) developed an assessment tool, the Strategic Alliance Formative Assessment Rubric, based on the aforementioned principles of collaboration. The tool can be used to help partnerships measure the relative strength of their partnership over time. Hays et al. (2000) also developed a measure of collaboration. Members were asked how frequently they engaged in six increasingly complex collaborative activities with other partners. Responses were measured on a 5-point Likert scale.

Researchers at the Amherst H. Wilder Foundation in St. Paul developed the Wilder Collaboration Factors Inventory that assesses the partnership's strengths and weaknesses (Mattessich, Murray-Close, & Monsey, 2001). Gardner (2000)

created instruments to examine collaborative values in California partnerships for substance abuse prevention.

The National Network for Collaboration (Borden & Perkins, 1999) developed a collaboration progress chart. The chart allows partnership members to rate the partnership on the following factors: goals, communicating, sustainability, research and evaluation, political climate, resources, catalysts, policies and reputations, history, connectedness, leadership, community development, and understanding the community. A definition of each of these factors is part of the instrument.

The Cooperative State Research, Education, and Extension Service, U.S. Department of Agriculture created five national networks to marshal faculty and program resources to respond to the economic, social, and human stresses faced by children, youth, and families (Bergstrom et al., 1995). These national networks created a collaboration framework to address community capacity. The framework is designed as a planning tool to develop and sustain collaboration as well as a diagnostic tool to evaluate ongoing development and progress.

Butterfoss, Goodman, Wandersman, Valois, and Chinman (1996b) developed a 129-item self-administered survey to measure the effectiveness of committees in partnerships. The instrument was derived from existing instruments and tested for reliability (all but one had high internal consistency). Characteristics covered by the survey were leadership roles, staff-committee relationships, organizational climate, decision-making processes, community linkages, member satisfaction, member participation patterns, and members' costs and benefits. Van Houten, Castillo, Crompton, and Nobles (2000) developed a number of instruments to assess the establishment and effectiveness of networks. Goodman et al. (1996) developed a meeting effectiveness inventory asking respondents to rate the meeting's agenda, leadership, decision making, and value.

## Research Activities and Outcomes

Even though this article does not focus on measuring the long-term outcomes of partnerships, it is relevant to briefly discuss some of the challenges associated with such measurement and evaluation. It is precisely these challenges that make it important to document and measure the establishment and working of partnerships and their immediate effect on the capacity of all participants to address the targeted issues. In a report to congressional committees, the U.S. General Accounting Office (2003) stated that having collaborative partnerships is one of the key indicators of evaluation capacity.

Evaluations of the long-term effect of partnerships on targeted issues show a mixed record. Birkby (2003) reviewed literature on effectiveness of coalitions and identified a number of studies that were not able to conclusively demonstrate effectiveness of major initiatives. An evaluation of the Robert Wood Johnson Fighting Back program targeting drug use concluded, "Coalitions are expensive to



maintain and may not lend themselves to effective or well-implemented strategies” (Halfors et al., as cited in Birkby, 2003). Yin, Kaftarian, Yu, and Jansen (1997) evaluated the CSAP Community Partnership Program and found that only eight of the 24 communities studied showed statistically significant results lower than comparison communities on at least one of six outcomes examined.

On the other hand, Birkby’s (2003) review of the literature did identify a number of successful partnership collaborations. Berkowitz (2001), as reported by Birkby, found that they have achieved positive outcomes in the following areas: disability advocacy, education, health clinics, access to prenatal care, housing for the mentally ill, and physical exercise.

Wandersman and Florin (2003) identified successful outcomes for an arson prevention coalition in Detroit and the Consortium for the Immunization of Norfolk’s Children initiative in Norfolk, Virginia. As discussed by Wandersman and Florin, review the results of studies examining the effect of collaborative efforts targeting substance abuse and find that collaborative strategies targeting policy change appeared to be the most effective.

Birkby (2003) identified the following reasons why it may be so difficult to evaluate the long-term effectiveness of partnerships and coalitions:

- *Coalitions are not well defined.* Unique characteristics of each coalition make it difficult to replicate the initiative.
- *Extraneous variables can influence outcomes.* Moreover, extraneous variables differ from community to community. They include policy changes, new government initiatives, and population changes. All of these can interact with each other as well as with the community initiative.
- *It is difficult to match the community with the partnership initiative with a similar community without such an initiative.* Without such comparisons, however, it is difficult to attribute changes to the partnership.
- *It is difficult to draw conclusions across coalitions.* They often differ in intended outcomes, or worse yet do not have the same access to good baseline data.
- *The long-term effects may not be measurable until years later.* Many coalitions either do not measure intermediate outcomes or do not have well-articulated theory to link intermediate and long-term outcomes.
- *Coalitions may change in essential components.* Political pressure or pressure from funding sources may change the coalition’s structure or functioning.
- *Coalitions are multilayered and complex.* The complex nature of the coalition does not lend itself well to traditional evaluations.

Kaftarian and Yin (1997) discussed the methodological challenges of evaluating the outcomes of community-based partnerships, specifically partnerships for substance abuse prevention. When interventions target individuals, it may be

possible to randomly assign some individuals to the interventions and the others to a control group. This, however, is not feasible when the intervention targets an entire community system: its institutions, norms, behaviors, attitudes, and policies. In the latter case, the community itself is the unit of analysis, not the individuals; in these instances, individuals when studied are seen as subunits, nested within the overall unit of analysis. Furthermore, the open systems nature of the partnerships and the complex nature of communities make it very difficult to ascribe change. In a special journal edition on this topic, Kaftarian and Yin presented several approaches used to overcome these challenges. Although none of these were, or could be, experimental or quasi-experimental designs, they did each explore alternate explanations (or rival hypotheses) for the observed changes. Two of these methods included cross-community analysis in which the partnership community was matched with another community with similar demographic characteristics (Yin et al., 1997).

Kubisch, Fulbright-Anderson, and Connell (1998) described features of comprehensive community initiatives for children and families that make them difficult to evaluate:

- *Horizontal complexity.* They work across multiple sectors (social, economic, physical, political, and others) simultaneously and aim for synergy among them.
- *Vertical complexity.* They aim for change at the individual, family, community, organizational, and systems levels.
- *Community building.* They aim for strengthened community capacity, enhanced social capital, an empowered neighborhood, and similar outcomes.
- *Contextual issues.* They aim to incorporate external political, economic, and other conditions into their framework, even though they may have little power to affect them.
- *Community responsiveness and flexibility over time.* They are designed to be community specific and to evolve in response to the dynamics of the neighborhood and the lessons being learned by the initiative.
- *Community saturation.* They aim to reach all members of a community, and therefore individual residents cannot be randomly assigned to treatment and control groups for the purposes of assessing the [comprehensive community initiative] impact; finding equivalent comparison communities is also not feasible.

In a similar vein, Wandersman and Florin (2003) found fewer than expected community interventions (including but not limited to partnerships) that show the desired results. They recommend that future initiatives include “greater articulation of theory, increased sensitivity or measures, improved (or different) methods or designs, and expanded use of best practices.”

## Contextual Conditions, Capacity, and Rival Explanations

### *Partner Capacity*

Many of the measures mentioned elsewhere in this article, if administered at different points in time, can be used to measure increased capacity of the partnership and its partner organizations. Rothwell (2004) developed a set of self-assessment questions to measure institutional learning.

### *Community Capacity*

Bartle (2003) proposed analyzing the strength, power, or capacity of a community by measuring change in the following features of the community: altruism, common values, communal services, communications, confidence, political and administrative context, information, intervention, leadership, networking, organization, political powers, skills, and wealth. He recommended that community members (not just those in power) assess whether there has been an increase in any of these dimensions. However, to prevent bias, he recommended the collection of complementary data (such as the number and type of communal services). This includes facilitator handouts designed for participatory measurement of the strength of each of the aforementioned dimensions. The first measure provides an estimate of strength. Both measures examine the current status and ask participants for a retrospective assessment of change over the past 12 months and the previous five years.

Gardner (1995a) developed a collaborative assessment of capacity. The instrument is designed as a guide for county-level youth and family collaboratives. It covers 10 elements of collaborative capacity: governance and accountability, outcomes, financing, nonfinancial resources, community and parent ownership, staff and leadership development, program strategies, policy agenda development, organizational coherence, and addressing the equity issue.

Putnam (as cited in Hudson & Chapman, 2002) proposed a social capital questionnaire as a supplement to the 2002 Census Bureau's Current Population Survey. Grootaert, Narayan, Nyan, and Woolcock (2003) developed an instrument to measure social capital of communities in underdeveloped countries. Nevertheless, with some revision, some of the questions may be applicable to the MSPs. As Grootaert et al. pointed out, the content and phrasing of questions will not be appropriate in all countries, and locally important questions may need to be added. The Social Capital Questionnaire collects data on six dimensions: groups and networks, trust and solidarity, collective action and cooperation, information and communication, social cohesion and inclusion, and empowerment and political action.

Healy (2003) reviewed the international literature to identify measures of social capital (quite a few instruments exist to examine social capital in developing

countries). He concluded that “a single measure approach to social capital based on, for example, numbers of associations, membership rates or generalized trust offers a very limited means for measuring the extent of social capital.” He included examples and selections of questions on social capital from a number of international surveys. He recommended that the measurement of social capital be approached at a number of levels:

- Standardized questions on trust, civic engagement, social support networks, and so on, in large-scale household surveys
- Surveys of observed or reported human behavior
- Specific and contextual questions on relationships, attitudes, and behavior in community or organizational-specific surveys neighborhood, enterprise, or school
- Case-study, qualitative, or action-based research, which seeks to explore the meaning and interpretation of social interaction in a particular situation or context
- Randomized social experiments that seek to combine measurement with active policy intervention and “laboratory-simulated” conditions.

Bjornskov and Svendsen (2003) examined existing measurement systems and identified four dominant operational features of social capital measures: (a) the trust radius of a population as measured by the percentage of a population believing that people can be trusted; (b) the density of voluntary organizations in a given area, as measured by the number of organizations in which an average resident participates; (c) community members’ perceptions of honesty and corruption; and (d) measures of economic freedom. They concluded that one may need to divide social capital into two dimensions: one dimension in which social capital refers to honesty and trust in both individuals and institutions, and another dimension that refers to civic participation.

Gouvis and Moore (2004) used the following data sources to measure social capital in several District of Columbia neighborhoods: secondary data on organizations in the community, including the National Center for Charitable Statistics database (<http://nccs.urban.org>), and interviews with representatives of community organizations. Bullen and Onyx (1998) presented a social capital instrument and practitioners guide used to measure social capital in five communities in New South Wales, Australia.

A community organizational assessment tool developed by the Citizens Involvement Training Program at the University of Massachusetts–Amherst, as a mechanism to facilitate organizational discussion and development, may be relevant to partnership development (Bright, 1998). The Nonprofit Management Education Center of the University of Wisconsin Extension has developed a Strategic Alliances Assessment Tool that may be relevant to assessment of strategic planning



by partnerships (Lewis, 1998). The tool is based in part on the just-referenced community organizational assessment tool (Bright, 1998) and a checklist of nonprofit indicators developed in 1998 by the United Way of Minneapolis Area.

Hays et al. (2000) studied the relationship between the structure of substance abuse prevention coalitions and community impact. Next are the measures they used to assess the following constructs: sectorial representation and member diversity.

- *Sectorial Representation.* The members of each of 28 Illinois coalitions were asked to identify the community sector they represented from among 17 different sectors. Sectorial representation was measured as the total number of unique community sectors represented on a given coalition.
- *Member Diversity.* On the assumption that diversity usually means the inclusion of non-White members, member diversity was measured as the percentage of non-White members in a coalition.

Harms, Hines, Arnold, and Papsdorf (2001) developed a community readiness instrument and a sustainability assessment worksheet for Washington State's partnership for children's oral health. Bell-Elkins (2002) developed an instrument to assess principles of partnership in a community-campus partnership.

## CONCLUSION

Interorganizational partnerships play a critical role in research. Therefore, it is essential that partnerships make evaluation a priority. The MSP Program's partnerships are complex in structure and functioning, and developing a self-evaluation instrument would require considerable expertise and cost. Remarkably, a number of instruments do exist for effectively assessing and evaluating various components of partnerships; rather than developing a new instrument, partnerships should consider modification, use, and local validation of existing instruments, which provide reliable and short-term solutions.<sup>6</sup>

## ACKNOWLEDGMENT

This study is one in a series of substudies for the Math and Science Partnership Program Evaluation (MSP-PE) conducted for the National Science Foundation's MSP Program. The MSP-PE is conducted under Contract No. EHR-0456995.

---

<sup>6</sup>This article does not address issues related to how one selects instruments that might be adopted for local use or how to validate instruments for local use. General references for these tasks are included, however.

Since 2007, Bernice Anderson, Ed.D., Senior Advisor for Evaluation, Directorate for Education and Human Resources, has served as the National Science Foundation Program Officer. The author, from COSMOS Corporation, is Jennifer Scherer.

The MSP-PE is led by COSMOS Corporation in current partnership with George Mason University (GMU) and Brown University. Robert K. Yin (COSMOS) serves as Principal Investigator and Jennifer Scherer (COSMOS) serves as one of three Co-Principal Investigators. Additional Co-Principal Investigators and their collaborating institutions (including discipline departments and math centers) are Patricia Moyer-Packenham (USU, formerly GMU) and Kenneth Wong (Brown).

## REFERENCES

- Barnes, D., Carpenter, J., & Bailey, D. (2000). Partnership with service users in interprofessional education for community mental health: A case study. *Journal of Interprofessional Care, 14*(2).
- Bartle, P. (2003). *Measuring the strengthening of communities for community development participatory methods of measuring empowerment*. Unpublished manuscript.
- Bell-Elkins, J. (2002). *Case study of a successful community-campus partnership: Changing the environment through collaboration, doctoral dissertation*. Unpublished doctoral dissertation, University of Massachusetts, Boston.
- Bergstrom, A. R., Clark, T., Hogue, T., Iyechad, J., Miller, J., Mullen, S., et al. (1995). *Collaboration framework: Addressing community capacity*. Paper presented at the meeting of the National Network for Collaboration, Fargo, ND.
- Berkowitz, B. (2001). Studying the outcomes of community-based coalitions. *American Journal of Community Psychology, 29*(2), 213–227.
- Birkby, B. (2003). Community coalitions: Questions, controversy & context. *Prevention Evaluation Perspectives, 1*(1).
- Bjornskov, C., & Svendsen, G. T. (2003). *Measuring social capital: Is there a single underlying explanation?* (Working Paper No. 03-5). Aarhus, Denmark: Aarhus School of Business. Department of Economics.
- Borden, L., & Perkins, D. (1999). Assessing your collaboration: A self-evaluation tool. *Journal of Extension, 37*(2).
- Bright, R. (1998). *Community organization assessment tool* (Family Living Programs). Madison: University of Wisconsin–Extension. Retrieved from [http://www.uwex.edu/ces/cced/documents/assessment\\_tool\\_coat.pdf](http://www.uwex.edu/ces/cced/documents/assessment_tool_coat.pdf)
- Bullen, P., & Onyx, J. (1998). *Measuring social capital in five communities in NSW*. New South Wales, Australia: Management Alternatives for Social Services.
- Butterfoss, F. D., Goodman, R. M., & Wandersman, A. (1996a). Community coalitions for prevention and health promotion: Factors predicting satisfaction, participation and planning. *Health Education Quarterly, 23*(1), 6579.
- Butterfoss, F. D., Goodman, R. M., Wandersman, A., Valois, R. F., & Chinman, M. (1996b). The plan quality index: An empowerment evaluation tool for measuring and improving the quality of plans. In D. M. Fetterman, S. J. Kaftarian, & A. Wandersman (Eds.), *Empowerment evaluation: Knowledge and tools for self-assessment & accountability*. Thousand Oaks, CA: Sage.
- Callan, P. M. (1998). The role of state policy systems in fostering separation or collaboration. In P. M. Timpane & L. S. White (Eds.), *Higher education and school reform* (pp. 41–56). San Francisco, CA: Jossey-Bass.

- Clark, R. W. (1999). School-university partnerships and professional development schools. *Peabody Journal of Education*, 74, 164–177.
- Clegg, S., & McNulty, K. (2002). Partnership working in delivering social inclusion: Organizational and gender dynamics. *Journal of Education Policy*, 17, 5.
- Clifford, G., & Guthrie, J. W. (1988). *ED school: A brief for professional education*. Chicago: The University of Chicago Press.
- Committee on SMTP. (2001). *Educating teachers of science, mathematics, and technology: New practices for the new millennium*. Washington, DC: National Academy Press.
- Conference Board of the Mathematical Sciences. (2001). The mathematical education of teachers: Issues on mathematics education. *American Mathematical Society*, 11.
- Drug Strategies. (2001). *Assessing community coalitions*. Washington, DC: Author. Retrieved from <http://www.drugstrategies.org/commcoal>
- Elmore, R. F. (2000). *Building a new structure for school leadership*. Washington, DC: The Albert Shanker Institute.
- Gajda, R. (2004). Utilizing collaboration theory to evaluate strategic alliances. *American Journal of Evaluation*, 25(1), 65–77.
- Gardner, S. (1995a). *The four tools: Local collaborative assessment of capacity*. Irvine, CA: The Children's Policy Group.
- Gardner, S. (1995b). *Questionnaire on collaborative values* (Center for Collaboration for Children). Fullerton: California State University.
- Gardner, S. (2000). *Collaborative values inventory, proposition 10 version*. Fullerton: California State University.
- Gillespie, N. (2003). *Measuring trust in working relationships: The behavioral trust inventory*. Victoria, Australia: University of Melbourne.
- Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *The Quarterly Journal of Economics*, 115(3), 811–846.
- Goodlad, J. I. (1993). School-university partnerships and partner schools. *Educational Policy*, 7, 24–39.
- Goodlad, J. I., & Sirotnik, K. A. (1988). The future of school-university partnerships. In K. A. Sirotnik & J. I. Goodlad (Eds.), *School-university partnerships in action*. (pp. 205–225). New York: Teachers College.
- Goodman, R. M., Wandersman, A., Chinman, A., Imm, P., & Morrissey, E. (1996). An ecological assessment of community-based interventions for prevention and health promotion: Approaches to measuring community coalitions. *American Journal of Community Psychology*, 24(1), 33–61.
- Gouvis, R. C., & Moore, G. (2004). Measuring local institutions and organizations: The role of community institutional capacity in social capital. *Measuring Local Institutions and Organizations* [Abstract]. Retrieved from <http://www.urban.org/Template.cfm?NavMenuID=24&template=/TaggedContent/ViewPublication.cfm&PublicationID=8852>
- Grootaert, C., Narayan, D., Nyan, J. V., & Woolcock, M. (2003). *Integrated questionnaire for the measurement of social capital*. Washington, DC: The World Bank.
- Harms, R., Hines, B., Arnold, T., & Papsdorf, B. (2001). *Community roots for oral health: Guidelines for successful coalitions* [Brochure]. Olympia, WA.
- Haycock, K., Hart, P., & Irvine, J. (Eds.). (1992). *Improving student achievement through partnerships*. Alexandria, VA: American Association of Higher Education.
- Hays, C. E., Hays, S. P., DeVille, J. O., & Mulhall, P. F. (2000). Capacity for effectiveness: The relationship between coalition structure and community impact. *Evaluation and Program Planning*, 23, 373–379.
- Healy, T. (2003, March). *Social capital: Challenges for its measurement at international level*. Paper presented at the Anniversary Conference on Sustainable Ties in the Information Society.
- Holmes Group. (1990). *Tomorrow's schools: Principles for the design of professional development schools*. East Lansing, MI: Author.

- Howard Community College. (1999). *Collaboration with other education organizations*. Columbia, MD: Author. Retrieved from <http://www.howardcc.edu/future/edpart.pdf>
- Hudson, L. & Chapman, C. (2002, September). *The measurement of social capital in the United States*. Paper presented at the International Conference on the Measurement of Social Capital, London, England.
- Kaftarian, S. J. & Yin, R. K. (1997). Introduction: Challenges of community-based program outcome evaluations. *Evaluation and Program Planning*, 20(3), 293–297.
- Kingsley, G., & O'Neil, D. (2004, March). *Performance measurement in public-private partnerships: Learning from praxis, constructing a conceptual model*. Paper presented at the American Society for Public Administration 65th National Conference, Portland, OR.
- Kubisch, A. C., Fulbright-Anderson, K., & Connell, J. (1998). *Evaluating community initiatives: A progress report* (The Aspen Institute Roundtable on Community Change). Washington, DC: Aspen Institute.
- Langenberg, D. H., Marx, G. L., & Shapiro, N. S. (1999). Coordinating efforts for systemwide change: K-16 partnerships. *High School Magazine*, 6, 8–13.
- Lantz, P. M., Viruell-Fuentes E., Israel, B. A., Softley, D., & Guzman, R. (2001). Can communities and academia work together on public health research? Evaluation results from a community-based participatory research partnership in Detroit. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, 78(3).
- Lewis, A. (1998). *Nonprofit organizational assessment tool* (Strategic Alliances). Madison: University of Wisconsin–Extension, Nonprofit Management Education Center.
- Marra, M. (2004). Knowledge partnerships for development: What challenges for evaluation? *Evaluation and Program Planning*, 27(2), 151–160.
- Mattessich, P., Murray-Close, M., & Monsey, B. (2001). *Collaboration what makes it work?* Minneapolis-St. Paul, MN: The Amherst H. Wilder Foundation.
- Metzler, M. M., Higgins, D. L., Beeker, C. G., Freudenberg, N., Lantz, P. M., & Senturia, K. D., et al. (2003). Addressing urban health in Detroit, New York City, and Seattle through community-based participatory research partnerships. *American Journal of Public Health*, 93(5), 803–811.
- Paine, K. (2003). *Guidelines for measuring trust in organizations*. Gainesville: The Institute for Public Relations. Commission on PR Measurement and Evaluation, University of Florida.
- Raizen, S. A., McLeod, D. B., & Howe, M. B. (1997). The changing conceptions of reform. In S. A. Raizen & E. D. Britton (Eds.), *Bold ventures: Patterns among innovations in science and mathematics education* (pp. 97–129). Dordrecht, the Netherlands: Kluwer Academic.
- Rice, E. H. (2002). The collaborative process in professional development schools: Results of a meta-ethnography, 1990-1998. *Journal of Teacher Education*, 53(1), 55–67.
- Rothwell, W. J. (2004). *The role of organizational learning in a learning organization*. New York: American Management Association.
- Sackett, K., Hendricks, C., & Pope, R. (2000). *Collaboration: An innovative education/business partnership*. New York: Buffalo School of Nursing.
- Schulz, A. J., Israel, B. A., & Lantz, P. (2003). Instrument for evaluating dimensions of group dynamics within community-based participatory research partnerships. *Evaluation and Program Planning*, 26(3), 249–262.
- Shah, J., & Singh, N. (2001). Benchmarking internal supply chain performance: Development of a framework, *The Journal of Supply Chain Management*, 37(1), 37–47.
- Teitel, L. (1993). The state role in jump-starting school/university collaboration: A case study. *Educational Policy*, 7, 74–95.
- The Annenberg Foundation. (2002). *The Annenberg challenge, lessons and reflections on public school reform*. Providence, RI: Annenberg Institute for School Reform. Available from [http://www.annenbergfoundation.org/other/other\\_show.htm?doc\\_id=212527](http://www.annenbergfoundation.org/other/other_show.htm?doc_id=212527)
- Tierney, W. G. (Ed.). (2001). *Faculty work in schools of education: Rethinking roles and rewards for the twenty-first century*. Albany: State University of New York Press.



- Timpane, M. P., & White, L. S. (Eds.). (1998). *Higher education and school reform*. San Francisco, CA: Jossey-Bass.
- Timperley, H., & Robinson, V. (2002). *Partnership: Focusing the relationship on the task of school improvement*. Wellington: New Zealand Council for Educational Research.
- U.S. Department of Education. (2003, December). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide*. Presented at the National Center for Education Evaluation and Regional Assistance, Washington, DC.
- U.S. General Accounting Office. (2003, May). *Program evaluation: An evaluation culture and collaborative partnerships help build agency capacity*. Report to Congressional Committees, Washington, DC.
- Van Houten, T., Castillo, I., Crompton, D., & Nobles, K. (2000). *Infrastructure and capacity to address issues of health disparity. Final report*. Bethesda, MD: Office of Minority Health, U.S. Department of Health and Human Services, COSMOS Corporation.
- Vinten, G. (1996). The education-business partnership: An evaluation. *The International Journal of Education Management*, 10(2), 27–38.
- Wandersman, A., & Florin, P. (2003). Community interventions and effective prevention. *American Psychologist*, 58(6/7), 441–448.
- Warrick, L. H., Wood, A. H., Meister, J. S., & de Zapien, J. G. (1992). Evaluation of a peer health worker prenatal outreach and education program for hispanic farmworker families. *Journal of Community Health*, 17(1), 13–26.
- Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 21, 1–19.
- Yin, R. K., Kafterian, S. J., Yu, P., & Jansen, M. A. (1997). Outcomes from CSAP's community partnership program: Findings from the national cross-site evaluation. *Evaluation and Program Planning*, 20, 345–355.

## APPENDIX

### Additional Sources

- Ariño, A., de la Torre, J., & Smith, P. R. (2001). The role of trust in inter-organizational alliances: Relational quality and partner behavior. *California Management Review*. See also: <http://www.anderson.ucla.edu/research/ciber/wponline/pdf/99-22.pdf>
- The Aspen Institute. (n.d.). *Measures for community research* (Roundtable for Comprehensive Community Initiatives). See also: <http://www.aspenmeasures.org/index.html>
- Baum, H. S. (2001). How should we evaluate community initiatives? *Journal of the American Planning Association*, 67(2), 147–158.
- Boris, E. T. (1999). Nonprofit organizations in a democracy: Varied roles and responsibilities. In E. T. Boris & C. E. Steuerle (Eds.), *Nonprofits & government, collaboration and conflict*. Washington, DC: The Urban Institute.
- Bourdieu, P. (1983). Forms of social capital. In J. C. Richards (Ed.), *Handbook of theory and research for sociology of education* (pp. 241–258). New York: Greenwood.
- Brownson, R. C., Baker, E. A., & Kreuter, M. W. (2001). Prevention research partnerships in community settings: What are we learning? *Journal of Public Health Management and Practice*, 7(2), vii–ix.
- Butterfoss, F. D., Goodman, R. M., & Wandersman, A. (1993). Community coalitions for prevention and health promotion. *Health Education Research*, 8, 315–330.
- Carr, J. A. (2003). *It's not just academic: University–community partnerships are rebuilding neighborhoods. Housing facts and findings*. Washington, DC: Building Higher Education Community Development Corporation Partnerships, Office of University Partnerships, U.S. Department of Housing and Urban Development.

- Chaskin, R. J. (2001). Building community capacity: A definitional framework and case studies from a comprehensive community initiative. *Urban Affairs Review*, 36(3), 291–323.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94, 95–120.
- Connell, J. P., Kubisch, A. C., Schorr, L. B., & Weiss, C. H. (1995). *New approaches to evaluating community initiatives volume 1: Concepts, methods, and contexts*. Washington, DC: The Aspen Institute.
- Davis, D., & McCullough, J. (2006, February). *Toward a taxonomy of partnerships: Theory and practice*. Paper presented for the National Science Foundation Directorate for Education and Human Resources, Arlington, VA.
- Dugery, J., & Knowles, J. (n.d.). *What we learned from others in the field: Key findings from the roundtable*. Charlottesville, VA: Pew Partnership for Civic Change.
- Ellis, T. M., & Lenczner, S. J. (2000). *Lessons from the field: Community anti-drug coalitions as catalysts for change* (A report to the Annie E. Casey Foundation). Alexandria, VA: Community Anti-Drug Coalitions of America.
- Ferraiolo, K., & Freedman, P. (n.d.). *Solutions for America: A collaborative research model* (pp. 7–20). Charlottesville, VA: University + Community Research Partnerships.
- Florin, P., & Chavis, D. M. (1990). *Community development and substance abuse prevention* (p. 112). San Jose, CA: Prevention Office, Bureau of Drug Abuse Services, Department of Health.
- Florin, P., Mitchell, R., & Stevenson, J. (1993). Identifying training and technical assistance needs in community coalitions: A developmental approach. *Health Education Research*, 8(3), 417–432.
- Francisco, V. T., Paine, A. L., & Fawcett, S. B. (1993). A methodology for monitoring and evaluating community health coalitions. *Health Education Research*, 8(3), 403–416.
- Francisco, V. T., & Butterfoss, F. D. (2003). How do we know if we are making a difference with our program or community initiative? *Health Promotion Practice*, 4(4), 367–370.
- Gabriel, R. (2000). Methodological challenges in evaluating community partnerships and coalitions: Still crazy after all these tears. *Journal of Community Psychology*, 28, 339–352.
- Gardner, S. (n.d.). *The community-based organization self-assessment instrument*. Fullerton: California State University, Center for Collaboration for Children.
- Ghere, R. (2001). Probing the strategic intricacies of public-private partnership: The patent as a comparative reference. *Public Administration Review*, 61(4), 441.
- Goodlad, J. I. (1990). *Teachers for our nation's schools*. San Francisco, CA: Jossey-Bass.
- Goodman, R. M., Steckler, A., Hoover, S., & Schwartz, R. (1993). A critique of contemporary community health promotion approaches: Based on a qualitative review of six programs in Maine. *American Journal of Health Promotion*, 7(3), 208–221.
- Goodman, R. M., & Steckler, A. B. (1987–1988). The life and death of a health promotion program: An institutionalization case study. *International Quarterly of Community Health Education*, 1, 5–21.
- Gouvis-Roman, C., Moore, G., Jenkins, S., & Small, K. (2002). *Understanding community justice partnerships: Assessing the capacity to partner*. Washington, DC: The Urban Institute. See also: <http://www.urban.org/Template.cfm?NavMenuID=24&template=/TaggedContent/ViewPublication.cfm&PublicationID=8404>
- Greenberg, J. S., Howard, D., & Desmond, S. (2003). A community-campus partnership for health: The seat pleasant-university of Maryland health partnership. *Health Promotion Practice*, 4(4), 393–401.
- Harkavy, I. (1998, January). *School-community-university partnerships: Effectively integrating community building and education reform*. Paper presented at the Conference on Connecting Community Building and Education Reform: Effective School. Community University Partnerships, Washington, DC. See also <http://www.gwu.edu/~bsep/255/community.htm>
- Harper, R. (2002, September). *The measurement of social capital in the United Kingdom*. Paper presented at the International Conference on the Measurement of Social Capital, London. See also <http://www.oecd.org/dataoecd/22/52/2382339.pdf>

- Harvard School of Public Health and the Global Health Council. (2000). *Public-private partnerships in public health*. Paper presented at the workshop proceedings, Dedham, Massachusetts.
- Healy, T. (2002, September). *The measurement of social capital at international level*. Paper presented at the International Conference on the Measurement of Social Capital, London.
- Herman, J. L., & Winters, L. (1992). *Tracking your school's success: A guide to sensible evaluation*. Newbury Park, CA: Corwin Press.
- Himmelman, A. T. (1996). On the theory and practice of transformational collaboration: From social service to social justice. In C. Huxham (Ed.), *Creating collaborative advantage*. (pp. 19–43). London: Sage.
- Hoel, J. (1998). *Cross-system collaboration: Tools that work*. Washington, DC: Child Welfare League of America. See also [http://www.cwla.org/programs/incarcerated/cop\\_getinvolved.htm](http://www.cwla.org/programs/incarcerated/cop_getinvolved.htm)
- Holland, B. (2001). *Characteristics of engaged institutions and sustainable partnerships, and effective strategies for change*. Retrieved from [http://www.servicelearning.org/resources/fact\\_sheets/he\\_facts/he\\_partners/](http://www.servicelearning.org/resources/fact_sheets/he_facts/he_partners/)
- Holland, G., Green, S., Lawrence, W., Green-Moton, E., & Stanton, T. K. (2003, April). *Community-university partnerships: What do we know?* Paper presented at the Community-University Partnerships: Translating Evidence into Action, San Diego, CA.
- Hord, S. M. (1986). A synthesis of research on organizational collaboration. *Educational Leadership*, 43, 22–26.
- Intriligator, B. A. (1992). *Establishing interorganizational structures that facilitate successful school partnerships*. San Francisco: American Educational Research Association. (ERIC Document Reproduction Service No. ED347692)
- Jasuja, K. G., Chou, P., Bernstein, K., Wang, E., McClure, M., & Pentz, M. (2005). Using structural characteristics of community coalitions to predict progress in adopting evidence-based prevention programs. *Evaluation and Program Planning*, 28, 173–184.
- Jehl, J., Blank, M., & McCloud, B. (2001, July/August). Lessons in collaboration: Bringing together educators and community builders. [Electronic version]. *Shelterforce Online* Retrieved from <http://www.nhi.org/online/issues/118/JehlBlankMcCloud.html>
- Kawachi, I. (2000). *Social capital*. The John D. and Catherine T. MacArthur Research Network on Socioeconomic Status and Health. See also <http://www.macses.ucsf.edu/Research/SocialEnvironment/notebook/capital.html> - Forms
- Kingsley, G. (2005). *Frames of reference for partnerships and some preliminary findings from an e-delphi on partnerships*. Available from [http://hub.mspnet.org/index.cfm/11110/post\\_show/post\\_id-456?listsorton=2&listsortdir=a](http://hub.mspnet.org/index.cfm/11110/post_show/post_id-456?listsorton=2&listsortdir=a)
- Kettl, D. F. (1988). *Government by proxy: (Mis?)managing federal programs*. Washington, DC: CQ Press.
- Kettl, D. F. (1993). *Sharing power: Public governance and private markets*. Washington, DC: The Brookings Institute.
- Kreuter, M. W., Lezin, N. A., & Young, L. A. (2000). Evaluating community-based collaborative mechanisms: Implications for practitioners. *Health Promotion Practice*, 1, 49–63.
- Kumpfer, K. L. & Chavez, N. R. (2000). *Prevention works through community partnerships*. Rockville, MD: SAMHSA's National Clearinghouse on Alcohol and Drug Information.
- Linder, S. H. (2000). Coming to terms with the public-private partnership: A grammar of multiple meanings. In P. V. Roseanau (Ed.), *Public-private policy partnerships*. London: MIT Press.
- Lewis, J. (2000). *Build trust trusted partners: How companies build mutual trust and win together*. New York: Simon & Schuster (1999). See also: <http://www.simonsays.com/subs/excerpt.cfm?areaid=286&isbn=0684836513>
- M & P Henderson & Associates Pty Ltd. (2004). *Partnerships for service delivery: review of the research and practice literature*. Report to Department of Communities, Queensland Government,

- Brisbane, Australia. See also: [http://www.getinvolved.qld.gov.au/share\\_your\\_knowledge/resources/documents/pdf/partnershipreview.pdf](http://www.getinvolved.qld.gov.au/share_your_knowledge/resources/documents/pdf/partnershipreview.pdf)
- Martinelli, F. (2001, September). *Growing and sustaining the partnership. How community-based organizations can start charter schools* (chap. 5). Retrieved from <http://www.uscharterschools.org/gb/community/5.htm>
- Mattessich, P. (2003, May/June). Can this collaboration be saved? Twenty factors that can make or break any group effort. *Shelterforce Online*, 129. Retrieved from <http://www.nhi.org/online/issues/129/savecollab.html>
- Mattessich, P., & Monsey, B. (1992). *Collaboration: What makes it work. A review of research literature on factors influencing successful collaboration*. Minneapolis-St. Paul, MN: The Amherst H. Wilder Foundation.
- Maurana, C., Beck, B., & Newton, L. (2001, June). *Community-campus partnership*. Paper presented at the National Advisory Committee Networking Meeting, Johns Hopkins Urban Health Institute, Baltimore, MD.
- McDonnell, L. M., & Elmore, R. F. (1987). Getting the job done: Alternative policy instruments. *Educational Evaluation and Policy Analysis*, 9(2), 133–152.
- Mitchell, R. P., Florin, P., & Stevenson, J. (2002). Supporting community-based prevention and health promotion initiatives: Developing effective technical assistance systems. *Health Education & Behavior*, 29(5), 620–639.
- Mosher, F. C. (1980). The changing responsibilities and tactics of the federal government. *Public Administration Review*, 40, 541–548.
- MSP-PE. (2005, September). *Evaluation design* (first annual report; Robert K. Yin, Kenneth K. Wong, Patricia S. Moyer-Packenham, and Jennifer Scherer). Submitted to the National Science Foundation.
- North, D. (1990). *Institutions, institutional change and economic performance*. New York: Cambridge University Press.
- Oliver, C. (1990). Determinants of interorganizational relationships: Integration and future directions. *Academy of Management Review*, 15(2), 241–265.
- Ouellette, P. M., Briscoe, R., & Tyson, C. (2004). Parent-school and community partnerships in children's mental health: Networking challenges, dilemmas, and solutions. *Journal of Child and Family Studies*, 13(3), 295–308.
- Phillips, J., Rivo, M. L., & Talamonti, W. J. (2004). Partnership between health care organizations and medical schools in a rapidly changing environment: A view from the delivery system. *Family Medicine*, 36, S121–S125.
- Pimmel, R. (2004, October). *Partnerships for engineering education*. Paper presented to the ASEE North Midwest Regional Conference, National Science Foundation, Milwaukee, WI.
- Provan, K. G., & Milward, J. B. (2001). Do networks really work? A framework for evaluating public-sector organizational networks. *Public Administration Review*, 61(4), 414–423.
- Public Management Institute. (2003). *Measuring trust: Government initiatives*. Katholieke Universiteit Leuven, Leuven, Belgium. See also: <http://www.soc.kuleuven.ac.be/io/trust/governmentinitiatives.htm>
- Putnam, R. D. (1993). The prosperous community: Social capital and public life. *American Prospect*, 13, 35–42.
- Putnam, R. (1996). Bowling alone: America's declining social capital. *Journal of Democracy*, 6(1), 65–79.
- Rackham, N., Friedman, L., & Buff, R. (1996). *Getting partnering right: How market leaders are creating long-term advantage*. New York: McGraw-Hill.
- Reich, M. R. (Ed.). (2002). *Public-private partnerships for public health*. Cambridge, MA: Harvard University Press.



- Reiniger, A. (2003). *Building community campus partnerships in underserved communities: An essential step to effective services*. Seattle, WA: Case Study for the Principles of Partnership, Community Campus Partnerships for Health.
- The Robert Wood Johnson Foundation. (1998). *Grant results report: Developing a public health performance monitoring system*. Available from <http://www.rwjf.org/reports/grr/024336s.htm>
- Salamon, L. M. (1981). Rethinking public management: Third-party government and the changing forms of government action. *Public Policy*, 29, 255–275.
- Sandman, L., & Baker-Clark, C. (1997). *Characteristics and principles of university-community partnerships: A delphi study* [Abstract]. Paper presented at the Midwest Research-to-Practice Conference in Adult, Continuing and Community Education Conference, Michigan State University.
- Savas, E. S. (2000). *Privatization and public-private partnerships*. New York: Chatman House.
- Saxe, L., Reber, E., Hallfors, D., Kadushin, C., Jones, D., Rindskopf, D., et al. (1997). Think globally, act locally: Assessing the impact of community-based substance abuse prevention. *Evaluation and Program Planning*, 20(3), 357–366.
- Scantlebury, K., Boone, W., Kahle, J. B., & Fraser, B. J. (2001). Design, validation, and use of an evaluation instrument for monitoring systemic reform. *Journal of Research in Science Teaching*, 38(6), 646–662.
- Schultz, A. J., Israel, B. A., & Lantz, P. (2003). Instrument for evaluating dimensions of group dynamics within community-based participatory research partnerships. *Evaluation and Program Planning*, 26(3), 249–262.
- Seifer, S., & Krauel, P. (2003). *Developing and sustaining equitable community-based participatory research partnerships*. Milwaukee, WI: Community-Campus Partnerships for Health. See also <http://www.ccph.info/>
- Sharp, P., Higham, J., Yeomans, D., & Mills, D. D. (2002). *Working together: The independent/state school partnerships scheme*. Retrieved from the Post-14 Research Group Web site: <http://www.leeds.ac.uk/educol/documents/00002223.htm>
- Smock, K. (1999). Building effective partnerships: The process and structure of collaboration. *Shelterforce Online*, 105. See also <http://www.nhi.org/online/issues/105/smock.html>
- Straub, D. W. (1989). Validating instruments in MIS research. *MIS Quarterly*, 13(2), 147–169.
- Thompson, L. S., Story, M., & Butler, G. (2003). Use of a university-community collaboration model to frame issues and set an agenda for strengthening the community. *Health Promotion Practice*, 4(4), 385–392.
- U.S. Department of Housing and Urban Development. (2002a). *Building higher education-Community Development Corporation (CDC) partnerships* [Abstract]. Washington, DC: Office of University Partnerships.
- U.S. Department of Housing and Urban Development. (2002b). *Partnership lessons*. Washington, DC: Office of University Partnerships.
- U.S. Department of Housing and Urban Development. (2002c). *Role of colleges and universities in supporting and strengthening community development corporations*. Washington, DC: Office of University Partnerships.
- U.S. Department of Health and Human Services. (1995). *Forging partnerships to eliminate tuberculosis*. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Prevention Services, Division of Tuberculosis Elimination.
- U.S. Department of Health and Human Services. (2003). *Global health partnerships*. Atlanta, GA: Centers for Disease Control and Prevention. See also <http://www.cdc.gov/ogh/partnerships.htm>
- U.S. Department of Health and Human Services. (n.d.). *Protecting the nation's health in an era of globalization: CDC's global infectious disease strategy*. Atlanta, GA: Centers for Disease Control and Prevention, Priority Area 6, Public Health Training and Capacity Building. See also [http://www.cdc.gov/globalidplan/17-priority\\_6.htm](http://www.cdc.gov/globalidplan/17-priority_6.htm)

- U.S. Department of the Interior. (n.d.). *Partnerships*. Washington, DC: National Park Service. See also <http://www.nps.gov/partnerships/about.htm>
- U.S. Department of Justice. (2001). *Building drug-free communities. A planning guide*. Alexandria, VA: Office of Juvenile Justice and Delinquency Prevention, Community Anti-Drug Coalitions of America.
- U.S. General Accounting Office. (1998). *Performance measurement and evaluation: Definitions and relationships* [Brochure]. Washington, DC: Author.
- Walker, C., Boris, E. T., Jackson, M., & Scott-Melnik, S. T. (1999). *Community partnerships for cultural participation: Concepts, prospects, and challenges*. Washington, DC: Urban Institute.
- Wielawski, I. (2002). *The fighting back program*. Princeton, NJ: The Robert Wood Johnson Foundation. See also: <http://www.rwjf.org/reports/nreports/fightingback.htm>
- Wiley, D. (2000). *Guidelines of best practices for partnership between tertiary institutions in Africa, and their foreign partners*. East Lansing, MI: Michigan State University. See also <http://africa.msu.edu/BestPractices.htm>
- Wolff, T. (2001a). Community coalition building—contemporary practice and research: Introduction. *American Journal of Community Psychology*, 29(2), 165–172.
- Wolff, T. (2001b). A practitioner's guide to successful coalitions. *American Journal of Community Psychology*, 29(2), 173–191.
- Wolff, T., & Foster, D. (1997). Principles of success in building community coalitions. In G. Kaye & T. Wolff (Eds.), *From the ground up! A workbook on coalition building & community development*. Amherst, MA: AHEC Community Partners.
- Woods, D. R. (1995). *Problem-based learning: Resources to gain the most from PBL*. Ontario, Canada: Author. See also <http://www.chemeng.mcmaster.ca/pbl/app-e.pdf#search=%22how%20to%20select%20evaluation%20instruments%22>
- The World Bank. (2003). *A user's guide to poverty and social impact analysis* (in conjunction with Poverty Reduction Group and Social Development). Washington, DC: Author. See also <http://www.worldbank.org/psia>
- Yin, R. K., & Ware, A. J. (2000). Using outcome data to evaluate community drug prevention initiatives: Pushing the state of the art. *Journal of Community Psychology*, 28(3), 323–338.

# Initial Trends in MSP-Related Changes in Student Achievement With MIS Data

Dimitar M. Dimitrov  
*George Mason University*

This substudy in the evaluation design of the Math and Science Partnership (MSP) Program Evaluation investigates changes in student mathematics and science achievement across three school years, 2002–03, 2003–04, and 2004–05, for MSP-related schools using Management Information System data with the Annual K-12 District Survey. First, changes in percentages of students (at or above) proficient on state assessments in math and science were investigated by gender, ethnicity, special education, and students with limited English proficiency using schools for which data were available for all three years. The results indicated that MSP schools continued to show improvement in student math and science proficiency over the three-year period. Second, schools were examined by frequency and effect size of increase, decrease, or no change in student math and science proficiency from the “start” (2002–03) to the “end” (2004–05) of the period for this study. The schools with positive changes were in much higher numbers and higher mean effect size of change compared to schools with negative (or no) changes in student math and science proficiency. Third, the relationship between the schools’ targeted teacher participation in MSP-related activities over the entire period of three years and the student math and science proficiency at the “end” year of this period (2004–05) was also investigated. It was found that this relationship was positive and significant for the elementary and high schools, but there was no evidence for its significance at the middle school level.

This study analyzes data from the Math and Science Partnership Management Information System (MSP-MIS) initiated by National Science Foundation as a Web-based data collection system. The purpose of the MSP-MIS is, in part, to assess the overall implementation of the MSP Program and to monitor the progress of individual MSP grants. Such implementation and monitoring are complex affairs

because of the complexity of the MSP grants. The MSP-MIS data are self-reported at the school level. Each grant is a partnership, minimally involving a K-12 district and an institution of higher education. More often, however, multiple districts and multiple institutions of higher education are engaged in a single MSP grant. The MSP-MIS collects annual data from all grantees, based on multiple instruments. Our study used data from only one of the instruments, the Annual K-12 District (school-level) Survey for years 2002–03, 2003–04, and 2004–05.

This study examines student proficiency in mathematics and science for MSP-related schools in terms of changes across three years (2002–03, 2003–04, and 2004–05) and relationships with MSP-related variables. It should be noted that the analysis does not control for differentiating MSP interventions, as such information is not provided with the MSP-MIS. Addressed are the following three research questions:

- RQ1: What are the trends in mathematics and science proficiency changes across the entire three-year period for MSP-related schools that were previously investigated (Dimitrov, 2005) for such changes across the first two years (2002–03, 2003–04)?
- RQ2: What is the frequency distribution of MSP-related schools across categories of change (increase, decrease, or no change) in math and science proficiency and what is the mean effect size for the categories of significant change (increase or decrease) from the first (2002–03) to the third (2004–05) year?
- RQ3: What is the relationship between a school's targeted teacher participation in MSP-related activities over the three-year period and the school's success in math and science proficiency at the end year of this period (2004–05)?

The first research question was investigated for a relatively small cohort of schools with MSP-MIS data on student proficiency in math and science at all three years (see Table 1)—the same schools that were previously investigated for changes in student math and science proficiency across the first two years (Dimitrov, 2005). This sample was used to examine the changes in math and science proficiency for the same schools across all three years. The results related to this research question (see the Results section for details) are summarized in Tables 2 and 3, with graphical representations in Figures 1 and 2, respectively.

The second research question was addressed with MSP-MIS data on student proficiency in math and science available at both Year 1 (2002–03) and Year 3 (2004–05). Practically, these were the data for the same schools used with RQ1, but there were a couple of additional schools since Year 2 (2003–04) data were not used with RQ2. Specifically, schools needed to report only Year 1 and Year 3 data, and there was now one additional middle school for math as well as three additional elementary schools and two middle schools for science (see Table 4).



TABLE 1

Math and Science Partnership—Management Information System Data for Number of Schools, Number of Students Assessed, and Number of Students at or Above Proficient at State Assessments in Mathematics and Science Across School Years 2002–03, 2003–04, and 2004–05

	Mathematics			Science		
	Elementary Schools	Middle Schools	High Schools	Elementary Schools	Middle Schools	High Schools
<b>All students</b>						
2002–03	<i>n</i> = 10,410 pass = 6,473 85 schools (96 %of all schools)	<i>n</i> = 8,406 pass = 4,398 25 schools (90% of all schools)	<i>n</i> = 1,869 pass = 787 8 schools (80% of all schools)	<i>n</i> = 1,527 pass = 1,028 20 schools (84% of all schools)	<i>n</i> = 5,039 pass = 2,752 17 schools (90% of all schools)	<i>n</i> = 282 pass = 127 3 schools (60% of all schools)
2003–04	<i>n</i> = 9,811 pass = 6,807 85 schools (27% of all schools)	<i>n</i> = 8,328 pass = 4,734 25 schools (15% of all schools)	<i>n</i> = 1,922 pass = 1,129 8 schools (5% of all schools)	<i>n</i> = 1,611 pass = 1,184 20 schools (15% of all schools)	<i>n</i> = 5,234 pass = 2,851 17 schools (22% of all schools)	<i>n</i> = 267 pass = 113 3 schools (3% of all schools)
2004–05	<i>n</i> = 14,514 pass = 11,010 85 schools (17%of all schools)	<i>n</i> = 10,548 pass = 6,167 25 schools (10% of all schools)	<i>n</i> = 2,035 pass = 1,205 8 schools (3% of all schools)	<i>n</i> = 1,839 pass = 1,335 20 schools (12% of all schools)	<i>n</i> = 5,055 pass = 2,988 17 schools (12% of all schools)	<i>n</i> = 276 pass = 128 3 schools (2% of all schools)
<b>Male</b>						
2002–03	<i>n</i> = 5,323 pass = 3,364 (85 schools)	<i>n</i> = 4,148 pass = 2,146 (24 schools)	<i>n</i> = 827 pass = 362 (6 schools)	<i>n</i> = 730 pass = 503 (18 schools)	<i>n</i> = 2,607 pass = 1,450 (17 schools)	<i>n</i> = 146 pass = 63 (3 schools)
2003–04	<i>n</i> = 4,988 pass = 3,446 (85 schools)	<i>n</i> = 4,162 pass = 2,355 (25 schools)	<i>n</i> = 999 pass = 572 (8 schools)	<i>n</i> = 749 pass = 578 (19 schools)	<i>n</i> = 2,664 pass = 1,485 (17 schools)	<i>n</i> = 136 pass = 57 (3 schools)
2004–05	<i>n</i> = 7,463 pass = 5,638 (85 schools)	<i>n</i> = 5,187 pass = 3,114 (24 schools)	<i>n</i> = 946 pass = 591 (6 schools)	<i>n</i> = 951 pass = 686 (20 schools)	<i>n</i> = 2,474 pass = 1,497 (16 schools)	<i>n</i> = 53 pass = 49 (1 schools)
<b>Female</b>						
2002–03	<i>n</i> = 5,067 pass = 3,103 (85 schools)	<i>n</i> = 3,965 pass = 2,094 (24 schools)	<i>n</i> = 780 pass = 355 (6 schools)	<i>n</i> = 780 pass = 516 (20 schools)	<i>n</i> = 2,419 pass = 1,294 (17 schools)	<i>n</i> = 136 pass = 64 (3 schools)
2003–04	<i>n</i> = 4,819 pass = 3,359 (85 schools)	<i>n</i> = 4,165 pass = 2,379 (25 schools)	<i>n</i> = 921 pass = 557 (8 schools)	<i>n</i> = 854 pass = 602 (20 schools)	<i>n</i> = 2,569 pass = 1,366 (17 schools)	<i>n</i> = 131 pass = 56 (3 schools)
2004–05	<i>n</i> = 7,048 pass = 5,371 (85 schools)	<i>n</i> = 5,180 pass = 3,007 (24 schools)	<i>n</i> = 901 pass = 566 (6 schools)	<i>n</i> = 888 pass = 649 (20 schools)	<i>n</i> = 2,401 pass = 1,426 (19 schools)	<i>n</i> = 37 pass = 33 (1 school)
<b>White</b>						
2002–03	<i>n</i> = 5,061 pass = 3,916 (53 schools)	<i>n</i> = 4,429 pass = 3,097 (21 schools)	<i>n</i> = 514 pass = 278 (5 schools)	<i>n</i> = 683 pass = 585 (9 schools)	<i>n</i> = 2,613 pass = 2,108 (14 schools)	<i>n</i> = 231 pass = 115 (2 schools)

(Continued on next page)

TABLE 1

Math and Science Partnership–Management Information System Data for Number of Schools, Number of Students Assessed, and Number of Students at or Above Proficient at State Assessments in Mathematics and Science Across School Years 2002–03, 2003–04, and 2004–05 (Continued)

	Mathematics			Science		
	Elementary Schools	Middle Schools	High Schools	Elementary Schools	Middle Schools	High Schools
2003–04	<i>n</i> = 4,871 pass = 4,013 (52 schools)	<i>n</i> = 4,732 pass = 3,463 (22 schools)	<i>n</i> = 835 pass = 595 (7 schools)	<i>n</i> = 676 pass = 616 (9 schools)	<i>n</i> = 2,686 pass = 2,146 (14 schools)	<i>n</i> = 227 pass = 102 (2 schools)
2004–05	<i>n</i> = 6,571 pass = 5,716 (78 schools)	<i>n</i> = 6,636 pass = 4,906 (23 schools)	<i>n</i> = 796 pass = 606 (6 schools)	<i>n</i> = 684 pass = 640 (18 schools)	<i>n</i> = 2,682 pass = 2,232 (16 schools)	<i>n</i> = 86 pass = 78 (1 schools)
African American						
2002–03	<i>n</i> = 1,757 pass = 728 (58 schools)	<i>n</i> = 1,896 pass = 338 (22 schools)	<i>n</i> = 36 pass = 17 (3 schools)	<i>n</i> = 159 pass = 75 (13 schools)	<i>n</i> = 1,649 pass = 279 (16 schools)	<i>n</i> = 3 pass = 3 (1 schools)
2003–04	<i>n</i> = 1,663 pass = 807 (59 schools)	<i>n</i> = 1,775 pass = 351 (22 schools)	<i>n</i> = 48 pass = 22 (5 schools)	<i>n</i> = 177 pass = 91 (13 schools)	<i>n</i> = 1,766 pass = 332 (16 schools)	<i>n</i> = 2 pass = 2 (1 schools)
2004–05	<i>n</i> = 2547 pass = 1502 (78 schools)	<i>n</i> = 1,984 pass = 380 (23 schools)	<i>n</i> = 46 pass = 11 (4 schools)	<i>n</i> = 238 pass = 145 (19 schools)	<i>n</i> = 1,566 p = 356 (16 schools)	
Hispanic–Latino						
2002–03	<i>n</i> = 2,954 pass = 1,447 (66 schools)	<i>n</i> = 1,288 pass = 541 (18 schools)	<i>n</i> = 993 pass = 403 (3 schools)	<i>n</i> = 577 pass = 286 (19 schools)	<i>n</i> = 434 pass = 192 (14 schools)	
2003–04	<i>n</i> = 2,576 pass = 1,520 (63 schools)	<i>n</i> = 1,384 pass = 672 (19 schools)	<i>n</i> = 989 pass = 497 (3 schools)	<i>n</i> = 647 pass = 386 (20 schools)	<i>n</i> = 486 pass = 233 (15 schools)	
2004–05	<i>n</i> = 4,914 pass = 3,426 (84 schools)	<i>n</i> = 1,425 pass = 632 (20 schools)	<i>n</i> = 985 pass = 526 (3 schools)	<i>n</i> = 860 pass = 500 (19 schools)	<i>n</i> = 494 pass = 244 (12 schools)	
Asian						
2002–03	<i>n</i> = 70 pass = 59 (20 schools)	<i>n</i> = 69 pass = 50 (10 schools)	<i>n</i> = 11 pass = 7 (2 schools)	<i>n</i> = 39 pass = 33 (9 schools)	<i>n</i> = 53 pass = 47 (7 schools)	
2003–04	<i>n</i> = 78 pass = 68 (17 schools)	<i>n</i> = 55 pass = 43 (9 schools)	<i>n</i> = 7 pass = 7 (2 schools)	<i>n</i> = 52 pass = 47 (9 schools)	<i>n</i> = 43 pass = 33 (6 schools)	
2004–05	<i>n</i> = 227 pass = 198 (59 schools)	<i>n</i> = 131 pass = 102 (17 schools)	<i>n</i> = 5 pass = 4 (1 school)	<i>n</i> = 53 pass = 49 (16 schools)	<i>n</i> = 94 pass = 80 (13 schools)	

(Continued on next page)

TABLE 1

Math and Science Partnership—Management Information System Data for Number of Schools, Number of Students Assessed, and Number of Students at or Above Proficient at State Assessments in Mathematics and Science Across School Years 2002–03, 2003–04, and 2004–05 (*Continued*)

	Mathematics			Science		
	Elementary Schools	Middle Schools	High Schools	Elementary Schools	Middle Schools	High Schools
<b>Other</b>						
2002–03	<i>n</i> = 568 pass = 323 (68 schools)	<i>n</i> = 724 pass = 372 (22 schools)	<i>n</i> = 315 pass = 82 (5 schools)	<i>n</i> = 69 pass = 49 (12 schools)	<i>n</i> = 289 pass = 126 (17 schools)	<i>n</i> = 48 pass = 9 (1 schools)
2003–04	<i>n</i> = 623 pass = 399 (72 schools)	<i>n</i> = 382 pass = 205 (27 schools)	<i>n</i> = 43 pass = 8 (4 schools)	<i>n</i> = 59 pass = 44 (23 schools)	<i>n</i> = 253 pass = 107 (19 schools)	<i>n</i> = 38 pass = 9 (1 schools)
2004–05	<i>n</i> = 255 pass = 168 (45 schools)	<i>n</i> = 185 pass = 96 (16 schools)	<i>n</i> = 13 pass = 10 (4 schools)	<i>n</i> = 9 pass = 5 (8 schools)	<i>n</i> = 43 pass = 13 (13 schools)	<i>n</i> = 4 pass = 4 (1 schools)
<b>Special education students</b>						
2002–03	<i>n</i> = 1,179 pass = 380 (57 schools)	<i>n</i> = 708 pass = 110 (21 schools)	<i>n</i> = 135 pass = 29 (4 schools)	<i>n</i> = 148 pass = 43 (14 schools)	<i>n</i> = 741 pass = 187 (16 schools)	
2003–04	<i>n</i> = 1,291 pass = 561 (63 schools)	<i>n</i> = 838 pass = 119 (22 schools)	<i>n</i> = 84 pass = 11 (3 schools)	<i>n</i> = 145 pass = 49 (10 schools)	<i>n</i> = 714 pass = 200 (15 schools)	
2004–05	<i>n</i> = 2,125 pass = 1,146 (85 schools)	<i>n</i> = 1,194 pass = 210 (23 schools)	<i>n</i> = 166 pass = 70 (5 schools)	<i>n</i> = 266 pass = 131 (19 schools)	<i>n</i> = 770 pass = 264 (16 schools)	
<b>Limited English Proficiency students</b>						
2002–03	<i>n</i> = 707 pass = 209 (29 schools)	<i>n</i> = 125 pass = 32 (7 schools)	<i>n</i> = 21 pass = 3 (3 schools)	<i>n</i> = 140 pass = 23 (12 schools)	<i>n</i> = 39 pass = 3 (1 school)	
2003–04	<i>n</i> = 790 pass = 396 (34 schools)	<i>n</i> = 163 pass = 38 (12 schools)	<i>n</i> = 24 pass = 4 (4 schools)	<i>n</i> = 155 pass = 45 (10 schools)	<i>n</i> = 94 pass = 21 (7 schools)	
2004–05	<i>n</i> = 1,547 pass = 851 (75 schools)	<i>n</i> = 146 pass = 30 (19 schools)	<i>n</i> = 24 pass = 3 (3 schools)	<i>n</i> = 213 pass = 65 (18 schools)	<i>n</i> = 102 pass = 13 (13 schools)	

*Note.* All high school entries, the Other 2004–05 science elementary school entry, and the Limited English proficiency students 2002–03 mathematics middle schools entry contain insufficient (or lack of) data for the analysis in this study. *n* = number of students assessed; pass = number of students who “pass” (at or above proficient) the state assessment.

TABLE 2  
Changes in Percentage of Students at or Above Proficient on Mathematics State Assessments

Student Demographics	School Level	Year 1 – Year 2 2002–03 – 2003–04	Year 2 – Year 3 2003–04 – 2004–05	Year 1 – Year 3 2002–03 – 2004–05
All	Elementary	+7.2 [5.9 ÷ 8.5]	+6.5 [5.3 ÷ 7.6]	+13.7 [12.5 ÷ 14.8]
	Middle	+5.3 [3.0 ÷ 6.0]	+1.6 [0.2 ÷ 3.0]	+6.1 [4.7 ÷ 7.6]
	High	+16.6 [13.4 ÷ 19.8]	NSS	+17.1 [14.0 ÷ 20.2]
Gender				
Male	Elementary	+5.9 [4.1 ÷ 7.7]	+6.5 [4.9 ÷ 8.1]	+12.3 [10.7 ÷ 14.0]
	Middle	+4.8 [2.7 ÷ 7.0]	+3.5 [1.4 ÷ 5.5]	+8.3 [6.3 ÷ 10.3]
	High	+13.5 [8.9 ÷ 18.1]	+5.2 [0.9 ÷ 9.6]	+18.7 [14.0 ÷ 23.4]
Female	Elementary	+8.5 [6.6 ÷ 10.3]	+6.5 [4.9 ÷ 8.1]	+15.0 [13.3 ÷ 16.6]
	Middle	+4.3 [2.1 ÷ 6.5]	NSS	+5.2 [3.2 ÷ 7.3]
	High	+15.0 [10.2 ÷ 19.7]	NSS	+17.3 [12.5 ÷ 22.1]
Ethnicity				
White	Elementary	+5.0 [3.4 ÷ 6.6]	+4.6 [3.3 ÷ 5.9]	+9.6 [8.2 ÷ 11.0]
	Middle	+3.3 [1.4 ÷ 5.1]	NSS	+4.0 [2.3 ÷ 5.7]
	High	+17.2 [11.9 ÷ 22.4]	+4.9 [0.6 ÷ 9.2]	+22.0 [16.8 ÷ 27.2]
African American	Elementary	+7.1 [3.8 ÷ 10.4]	+10.4 [7.4 ÷ 13.5]	+17.5 [14.5 ÷ 20.6]
	Middle	NSS	NSS	NSS
	High	NSS	NSS	NSS
Hispanic/Latino	Elementary	+10.0 [7.4 ÷ 2.7]	+10.7 [8.5 ÷ 13.0]	+20.7 [18.5 ÷ 23.0]
	Middle	+6.6 [2.8 ÷ 10.3]	-4.2 [-7.9 ÷ -0.5]	
	High	+9.7 [5.3 ÷ 14.1]		+12.8 [8.4 ÷ 17.2]
Asian	Elementary	NSS	NSS	NSS
	Middle	NSS	NSS	NSS
	High			
Other/Mixed	Elementary	+7.2 [1.6 ÷ 12.7]	NSS	+9.0 [1.8 ÷ 16.3]
	Middle	NSS	NSS	NSS
	High	NSS		
Special education students	Elementary	+11.2 [7.4 ÷ 5.1]	+10.5 [7.0 ÷ 13.9]	+21.7 [18.2 ÷ 25.2]
	Middle	NSS	+3.4 [0.1 ÷ 6.6]	NSS
	High	NSS	+29.1 [16.8 ÷ 41.4]	+20.7 [10.0 ÷ 31.4]
Limited English Proficiency students	Elementary	+20.6 [15.625.5]	+4.9 [0.6 ÷ 9.2]	+25.4 [21.0 ÷ 29.9]
	Middle	NSS	NSS	NSS
	High	NSS	NSS	NSS

Note. 95% confidence interval [in brackets] is provided for statistically significant changes. The results in the rows for high school level must be interpreted with caution due to insufficient data. An empty cell indicates missing data. NSS = not statistically significant.

The third research question investigated the schools for which MSP-MIS data on targeted teacher participation and student proficiency in math and science at year 2004–05 were available (see Table 5). The idea behind RQ3 was to investigate the relationship between two variables: (a) the school’s targeted teacher participation



TABLE 3

Changes in Percentage of Students at or Above Proficient on Science State Assessments

Student Demographics	School Level	Year 1 – Year 2 2002–03 – 2003–04	Year 2 – Year 3 2003–04 – 2004–05	Year 1 – Year 3 2002–03 – 2004–05
All	Elementary	+6.2 [3.0 ÷ 9.4]	NSS	+5.3 [2.2 ÷ 8.4]
	Middle	NSS	+4.6 [2.7 ÷ 6.6]	+4.5 [2.6 ÷ 6.4]
	High	NSS	NSS	NSS
Gender Male	Elementary	+8.3 [3.7 ÷ 12.8]	–5.0 [–0.9 ÷ –9.2]	NSS
	Middle	NSS	+ 4.8 [2.1 ÷ 7.5]	+4.9[2.2 ÷ 7.6]
	High			
Female	Elementary	NSS	NSS	+6.9 [2.5 ÷ 11.3]
	Middle	NSS	+6.2 [3.5 ÷ 9.0]	+5.9 [3.1 ÷ 8.7]
	High			
Ethnicity White	Elementary	+5.5 [2.1 ÷ 8.9]	NSS	+7.9 [4.7 ÷ 11.2]
	Middle	NSS	+3.3 [1.3 ÷ 5.4]	+2.5 [0.5 ÷ 4.6]
	High			
African American	Elementary	NSS	NSS	+13.8 [3.8 ÷ 23.7]
	Middle	NSS	+3.9 [1.2 ÷ 6.7]	+5.8 [3.1 ÷ 8.6]
	High			
Hispanic/Latino	Elementary	+10.1 [4.5 ÷ 15.7]	NSS	+ 8.6 [3.3 ÷ 13.8]
	Middle	NSS	NSS	NSS
	High			
Asian	Elementary	NSS	NSS	NSS
	Middle	NSS	NSS	NSS
	High			
Other/Mixed	Elementary	NSS	NSS	NSS
	Middle	NSS	NSS	NSS
	High	NSS		
Special education students	Elementary	NSS	+15.5 [5.4 ÷ 25.5]	+20.2 [10.3 ÷ 30.1]
	Middle	NSS	+6.3 [1.6 ÷ 11.0]	+9.0 [4.4 ÷ 13.7]
	High			
Limited English Proficiency students	Elementary	+12.6 [3.0 ÷ 22.2]	NSS	+14.1 [4.9 ÷ 23.3]
	Middle	+14.6 [0.3 ÷ 29.0]	NSS	
	High			

*Note.* 95% confidence interval [in brackets] is provided for statistically significant changes. The three All and the one Other/Mixed Year 1 – Year 2 high school entries must be interpreted with caution because of insufficient data. An empty cell indicates missing data. NSS = not statistically significant.

in MSP-related activities “accumulated” over all three years, and (b) student math and science proficiency at the end of this period.

It is important also to emphasize that the research questions addressed in this article are deliberately part of a broader MSP-PE investigation of student achievement. In this sense, the results in this article complement (and to some

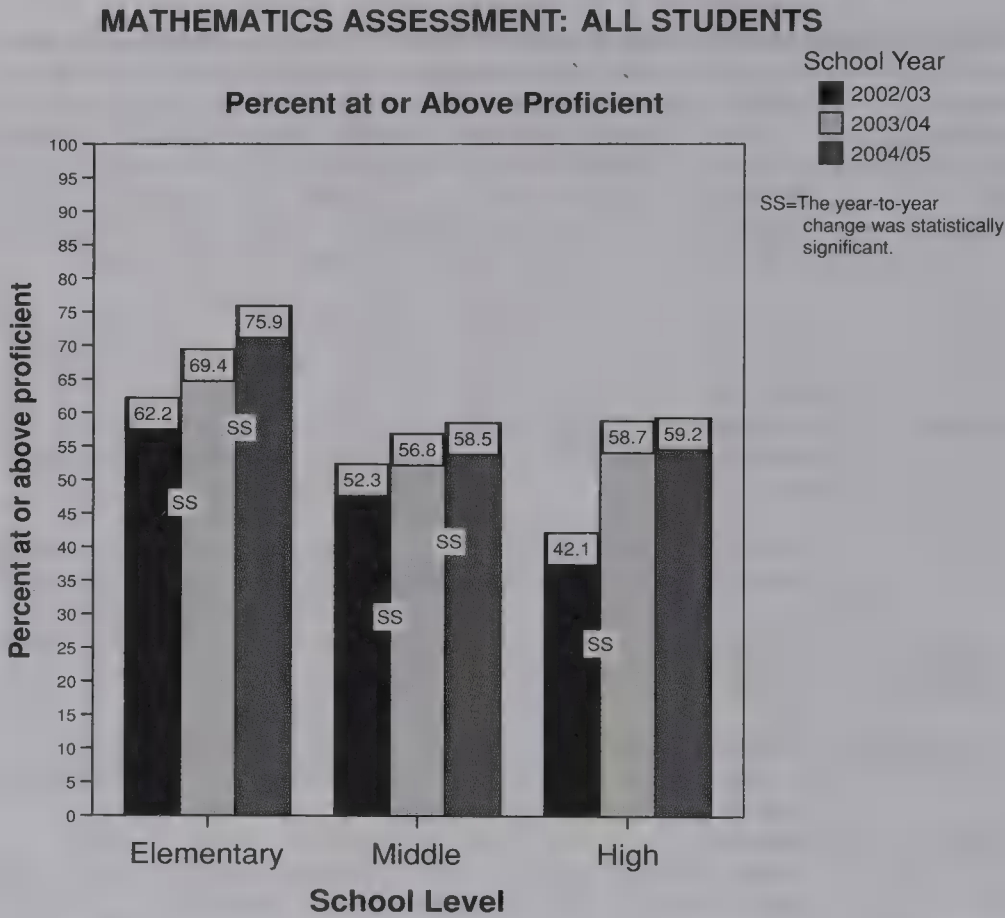


FIGURE 1 Percentage of students at or above proficient on mathematics state assessments.

degree “triangulate”) results reported in two other MSP-PE studies on student proficiency in math and science (Wong & Socha, this issue; Yin, in press).

Wong and Socha (2006) employed a broad and comprehensive analytic strategy in searching for a link between MSP activity and student math and science standardized test scores on a partnership-by-partnership basis. In their analysis, they use school-level state administrative data that are (in most cases) mandated to be reported by all public schools. This allows for a sizable control group of non-MSP participating schools for comparison purposes but is much more laborious in nature.

METHOD

Data

From the Annual K-12 District Survey, the data used in this article covered schools with available data for the three research questions as described in the

## SCIENCE ASSESSMENT: ALL STUDENTS

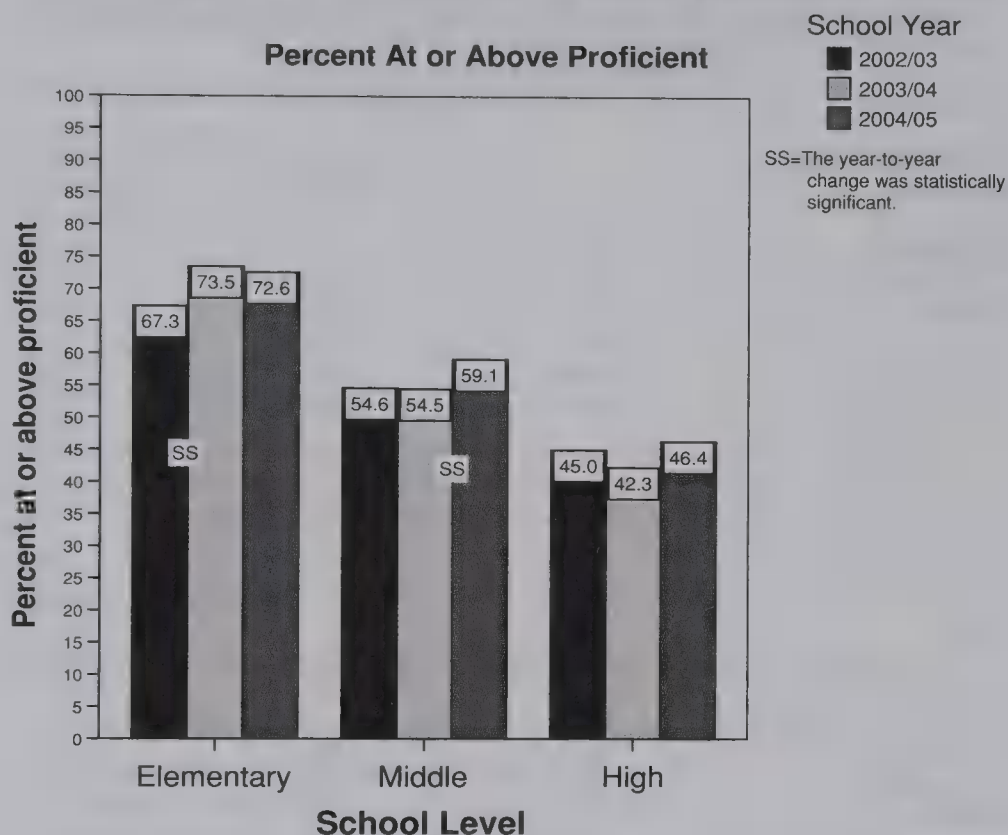


FIGURE 2 Percentage of students at or above proficient on science state assessments.

previous section. Table 1 provides (a) the number of schools for which MSP-MIS data on student math or science proficiency were available for all three years (2002–03, 2003–04, and 2004–05); (b) the percentage that these schools represent from all schools for which such data were available for the specific year; (c) the number of students in these schools who had taken the state assessment in math or science,  $n$ ; and (d) the number who “pass” (at or above proficient) the assessment. The data are also provided by gender, ethnicity, special education students, and limited English proficiency students. Table 1 shows, for example, that 85 elementary schools (10,410 students) represent 96% of all MSP-related schools for which MSP-MIS data on student math proficiency were reported for the first year, 2002–03. At the same time, these same schools represent only 27% and 17% of the MSP-related schools with such data for years 2003–04 and 2004–05, respectively. Table 1 also shows that the highest relative sample representation of schools is for mathematics at the elementary school level.

TABLE 4

Frequency of Schools With Direction and Mean Effect Size of Their Change From Year 2002–03 to 2004–05 in Percentages of Students at or Above Proficient on State Assessment in Mathematics and Science

Subject/School Level	Direction of Change		
	Increase	Decrease	No Change
<b>Mathematics</b>			
Elementary			
<i>N</i> (schools)	71	7	7
Effect size	0.45	0.18	—
<i>SE</i>	0.23	0.14	—
<i>n</i> (2002–03)	8709	895	806
<i>n</i> (2004–05)	12520	1107	887
Middle			
<i>N</i> (schools)	15	9	2
Effect size	0.20	0.19	—
<i>SE</i>	0.13	0.17	—
<i>n</i> (2002–03)	5179	3038	806
<i>n</i> (2004–05)	7078	3381	887
High			
<i>N</i> (schools)	5	3	0
Effect size	0.76	0.09	—
<i>SE</i>	0.44	0.10	—
<i>n</i> (2002–03)	853	2145	0
<i>n</i> (2004–05)	785	1710	0
<b>Science</b>			
Elementary			
<i>N</i> (schools)	16	4	3
Effect size	0.33	0.32	—
<i>SE</i>	0.16	0.13	—
<i>n</i> (2002–03)	1387	318	202
<i>n</i> (2004–05)	1576	390	220
Middle			
<i>N</i> (schools)	12	5	2
Effect size	0.15	0.09	—
<i>SE</i>	0.10	0.05	—
<i>n</i> (2002–03)	5179	3038	806
<i>n</i> (2004–05)	7078	3381	887
High			
<i>N</i> (schools)	2	1	0
Effect size	0.24	0.02	—
<i>SE</i>	0.19	0.00	—
<i>n</i> (2002–03)	1345	140	0
<i>n</i> (2004–05)	1197	146	0

*Note.* Effect size = Cohen's *h* (0.20 = small, 0.50 = medium, 0.80 = large). *n* = number of students during the school year (2002–03 or 2004–05). *SE* = standard error of the effect size across schools.



TABLE 5

Correlation Between School's Targeted Teacher Participation in MSP-Related Activities in Any of 3 Years (2002–03 to 2004–05) and School's Success in Student Achievement (Percentage of Students at or Above Proficient on State Assessments in Mathematics and Science) at the End Year (2004–05)

Subject/ School Level	<i>r</i>	<i>N</i>	<i>n</i>
Mathematics			
Elementary	.47**	128	30,272
Middle	.21	62	33,160
High	.53**	98	57,577
Science			
Elementary	.65**	46	4,700
Middle	.07	29	7,251
High	.38**	67	32,096

Note. *N* = number of schools (used for the calculation of the correlation coefficient, *r*); *n* = number of students who have taken the state assessment in these schools. \*\**p* < .01.

## Variables and Scales

There are three main variables investigated in this school-level study:

- *Student achievement*—The percentage of students at or above proficient on state assessments in mathematics and science.
- *Targeted teacher participation in MSP-related activities*—This variable is identified in MSP-MIS by the condition that 30% or more of a school's targeted teachers participated in 30 or more hr of MSP-sponsored activities during a single school year. Given the binary scale (1 if the condition was met, and 0 otherwise), the score for any school on this specific variable over three school years (2002–03, 2003–04, and 2004–05) may vary from zero to three (0 = the condition was not met during any of the three years, and 3 = the condition was met all three years).

## Statistical Analysis

For each school, the *changes* in student math and science achievement across three school years were measured by the differences in percentage of students at or above proficient on state assessments in mathematics and science as follows: Year 2 – Year 1, Year 3 – Year 2, and Year 3 – Year 1, where Year 1 = 2002–03, Year 2 = 2003–04, and Year 3 = 2004–05. A 95% confidence interval was calculated for each of these differences for each school. Changes in student mathematics and science achievement were also estimated by gender, ethnicity, special education students, and students with limited English proficiency. The Pearson

product-moment correlation was used to investigate the relationship between the school's targeted teacher participation in MSP-related activities over the period of all three years and student math and science proficiency at the end of this period.

## RESULTS

The results are reported in three parts representing the three research questions stated previously.

### Changes in Math and Science Proficiency Across Three Years for MSP-Related Schools

Figure 1 shows the percentage of students at or above proficient on state assessments in mathematics by school level (elementary, middle, and high schools). The statistically significant changes in this percentage (with 95% confidence intervals) are reported in Table 2. Similarly, the changes in percentage of students at or above proficient on state assessments in science are reported in Table 3 and graphically represented in Figure 2. Because the changes across the first two years (2002–03 to 2003–04) for the same schools were reported in a previous MSP-PE substudy (Dimitrov, 2005), the new information reported here shows the changes from Year 2 to Year 3 (2003–04 to 2004–05) and, more importantly, for the entire period from Year 1 to Year 3 (2002–03 to 2004–05).

The results in Tables 2 and 3, as well as their graphical representation in Figures 1 and 2, respectively, indicate a sustained trend of sizable positive changes in student math and science proficiency across the three years. One can expect that this trend is more stable for the elementary school level, given the relatively high number of schools (and students who had taken math and science assessments) at this school level (see Table 1). To what degree this trend holds for the middle and high schools remains to be examined with upcoming MSP-MIS data for subsequent years of school participation in MSP.

### Changes in Math and Science Proficiency From Year 2002–03 to Year 2004–05

To address the second research question, each school was tested for change in the proportion of math and science proficient students from Year 1 (2002–03) to Year 3 (2004–05). The school's performance was labeled (a) *increase*, if there was a statistically significant positive change; (b) *decrease*, if there was a statistically significant negative change; or (c) *no change*, if there was no statistically

significant change for the school. The Cohen's effect size index for a difference in two proportions,  $h$  (Cohen, 1988), was then calculated for each school with a statistically significant change (increase or decrease). Specifically, the  $h$  effect size for the difference in two proportions, say  $P_1 - P_2$ , is calculated as follows:

$$h = 2\arcsin\sqrt{P_1} - 2\arcsin\sqrt{P_2}$$

(Cohen, 1988, p. 181). The magnitude of the effect size is operationally defined as *small* ( $h = .20$ ), *medium* ( $h = .50$ ), and *large* ( $h = .80$ ; Cohen, 1988).

The frequency distribution of schools across the categories of increase, decrease, or no change in math and science proficiency is provided in Table 4. The mean effect size for elementary, middle, and high schools is also given in Table 4. The results indicate that the schools with a positive change (increase) clearly dominate in numbers and magnitude of mean effect size at the elementary, middle, and high school levels. For mathematics at the elementary school level, for example, there are 71 schools with a positive change (medium mean effect size of 0.45) versus seven schools with a negative change (small mean effect size of 0.18). It should be noted that the results for elementary schools in mathematics are more representative, compared to results for mathematics at middle and high school levels (or science at all school levels), given the larger number of elementary schools with MSP-MIS data on student math proficiency. Note also that RQ1 and RQ2 provide somewhat different information on the changes in student math and science proficiency from Year 1 (2002–03) to Year 3 (2004–05). Namely, although RQ1 relates to changes in percentage of students (at or above) proficient in math and science, RQ3 provides the frequency of schools across categories of change (increase, decrease, or no change) and the effect size of change with these categories.

### Targeted Teacher Participation in MSP-Related Activities and Student Proficiency

The third research question examines the relationship between the overall targeted teacher participation in MSP-related activities for any of three years and the student proficiency in math and science at the end of the three-year period (2004–05). The Pearson product-moment correlation coefficients for this relationship at the elementary, middle, and high school levels are provided in Table 5. The analysis shows a statistically significant positive relationship between the targeted teacher participation in MSP-related activities and student proficiency in math and science at the elementary and high school levels. However, there is no statistically significant relationship at the middle school level.

## DISCUSSION

This study examines initial trends in MSP-related changes in student math and science proficiency using MSP-MIS data with the Annual K-12 District Survey for three years, 2002–03, 2003–04, and 2004–05. The first research question addressed in this study relates to changes in the percentage of students (at or above) proficient in math and science for schools with available data on this variable for all three years. The second research question examines the changes in student math and science proficiency from the “start” (2002–03) to the “end” (2004–05) year in terms of frequency distribution of schools across categories of change in student math and science proficiency and the effect size of this change. The third research question examines the relationship between the school’s overall targeted teacher participation in MSP-related activities for all three years and the student proficiency in math and science at the “end” year, 2004–05.

### Changes in Student Math and Science Proficiency Across All 3 Years

The results on the first research question show that MSP-related schools maintain a trend of improvement in student math and science proficiency at all school levels across the three years. One can expect that this finding is particularly valid for the case of relatively large sample representation, math proficiency at the elementary school level, where the increase of 7.2% (at or above proficient) from Year 1 to Year 2 grows to 13.7% from Year 1 to Year 3 (see Table 2). This trend holds also for the groups defined in this study by gender, ethnicity, special education, and limited English proficiency. The lack of changes for the Asian students can be attributed to their relatively small sample representation and high performance on state assessments in math and science across all three years (see Table 1). It should be noted also that elementary, middle, and (particularly) high schools with data on science assessments are even further underrepresented compared to their counterparts with data on math assessments across all three years. It can be expected that follow-up MSP-PE studies, with much higher sample representation of MSP-related schools, will support the trend for sustained improvement in student math and science proficiency across all school levels.

### Changes in Math and Science Proficiency From Year 1 (2002–03) to Year 3 (2004–05)

The student sample size, the frequency distribution of schools across categories of change in student math and science proficiency, and the mean effect size for changes by school level are provided in Table 4. In addressing the second research



question, these results indicate that the schools with a positive change (increase) in student math and science proficiency clearly dominate in numbers and magnitude of mean effect size at the elementary, middle, and high school levels. For the most representative sample of schools (mathematics assessment for elementary schools), 84% of the schools show an increase, with a moderate mean effect size (Cohen's  $h = 0.45$ ); 6% of them show a decrease, with a small effect size (Cohen's  $h = 0.18$ ); and 6% of schools show no statistically significant overall change in student math proficiency from year 2002–03 to 2004–05.

### Targeted Teacher Participation in MSP-Related Activities and Student Proficiency

As for the third research question, the results show that there is a positive relationship between the school's targeted teacher participation in MSP-related activities, "accumulated" over the period of all three years, and the student proficiency in math and science at the elementary and high school levels (see Table 5).

### Limitations and Upcoming Analyses

The results in this study must be interpreted with an understanding of limitations that stem from restricted MIS data with the Annual K-12 District Survey. One limitation, for example, is the lack of matching data from "control" schools (not involved in MSP) to evaluate the degree to which the changes in students' proficiency in math and science can be attributed to school participation in MSP. That is why this study does not engage in testing hypotheses about the degree to which the delineated trends in math and science performance of MSP-related schools are different from trends that may exist in non-MSP related schools. Future triangulations with findings in other MSP-PE studies that control for MSP participation of schools (e.g., Wong & Socha, 2006) may provide more evidence on the role of MSP factors in the math and science proficiency of MSP-related schools.

Another limitation stems from the lack of MIS data that can be used to equate school proficiency measures in math and science across states. The purpose of such equating is to take into account differences (in content and passing standards) among state assessments in math and science. It should be noted, however, that mapping state performance standards on to a common scale (e.g., using National Assessment of Educational Progress data; Braun & Qian, 2007) is a difficult task still challenging the research on large-scale performance analyses (e.g., McLaughlin & Bandeira de Mello, 2003). From a different perspective, Yin, Schmidt, and Besag (2006) aggregated student achievement trends using standardized slopes as effect sizes across intervention sites and comparison sites.

In any case, the lack of a common scale for school (or student) performance of state assessments would be particularly damaging for results from comparisons across states. Such comparisons, however, are not targeted in this study. Instead, the focus is on changes in student math and science proficiency and its relationship with school's targeted teacher participation in MSP-related activities. Also, the binary measure of change in school proficiency (1 = statistically significant change and 0 otherwise), used with the second research question in this study, seems more "unified" across schools and thus more robust to aggregating compared to aggregating percentages of students (at or above) proficient in math or science state assessments at the school level.

In upcoming analyses with the continuation of this study, efforts will be directed to reducing validity threats associated with aggregation of student achievement trends across states—for example, through (a) mapping the aforementioned binary scores of change in school math or science proficiency on item response theory-derived scale, (b) weighting the proportions of students at or above proficient in math or science, (c) using standardized effect sizes, and (d) mapping state performance standards on to a common scale when appropriate data (collected outside MIS) is available. Additional analyses that can counteract the limitations with this study are also next steps in the MSP-PE agenda. Such analyses (e.g., using math and science course credit teacher training data) can further expand our understanding of the relationship between MSP-participation and student math and science achievement.

In conclusion, despite limitations in scope and depth of the analysis in this study, due primarily to data restrictions with the MIS Annual K-12 District Survey, the results indicate promising trends and relationships between student proficiency in mathematics and science and MSP-related variables.

## ACKNOWLEDGMENT

This study is one in a series of substudies for the Math and Science Partnership Program Evaluation (MSP-PE) conducted for the National Science Foundation's MSP Program. The MSP-PE is conducted under Contract No. EHR-0456995. Since 2007, Bernice Anderson, Ed.D., Senior Advisor for Evaluation, Directorate for Education and Human Resources, has served as the National Science Foundation Program Officer. The author from George Mason University, is Dimiter M. Dimitrov.

The MSP-PE is led by COSMOS Corporation in current partnership with George Mason University (GMU) and Brown University. Robert K. Yin (COSMOS) serves as Principal Investigator and Jennifer Scherer (COSMOS) serves as one of three Co-Principal Investigators. Additional Co-Principal Investigators and their collaborating institutions (including discipline departments and math centers) are Patricia Moyer-Packenham (USU, formerly, GMU) and Kenneth Wong (Brown).

## REFERENCES

- Braun, H., & Qian, J. (2007). An enhanced method for mapping state standards onto the NAEP scale. In N. J. Dorans, M. Pommerich, & W. P. Holland (Eds.), *Linking and aligning scores and scales* (pp. 313–338). New York: Springer.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Dimitrov, D. M. (2005). Analytic modeling with the MSP Management Information System. In *MSP-PE Evaluation Design and Drafts of Selected Substudies. First Annual Report, Vol. III* (pp. 1–26). Arlington, VA: National Science Foundation.
- McLaughlin, D., & Bandeira de Mello, V. (2003). *Comparing state reading and math performance standards using NAEP*. Paper presented at the National Conference on Large-Scale Assessment, San Antonio, CA.
- Yin, R. K. (in press). Student Achievement Data and Findings as Reported in MSPs' Annual and Evaluation Reports, *Journal of Education Research and Policy Studies*.
- Yin, R. K., Schmidt, R. J., & Besag, F. (2006). Aggregating student achievement trends across states with different tests: Using standardized slopes as effect sizes. *Peabody Journal of Education*, 81(2), 47–61.
- Wong, K. K., & Socha, T. (2006, November). *K-12 student achievement pilot study: Preliminary comparison of MSP and Non-MSP schools*. Preliminary report submitted to the National Science Foundation.
- Wong, K. K., & Socha, T. (2008/this issue). A pilot study to identify comparison schools for math and science partnership participating schools: Preliminary findings on one math/science partnership. *Peabody Journal of Education*, 83, 654–673.

# A Pilot Study to Identify Comparison Schools for Math and Science Partnership Participating Schools: Preliminary Findings on One Math/Science Partnership

Kenneth K. Wong  
*Brown University*

Ted Socha  
*National Center for Education Statistics*

This pilot study proposes a set of analytical steps for comparing schools that participate in the National Science Foundation's Math and Science Partnership (MSP) Program and their nonparticipating peers in the same state. This pilot is part of a larger effort to evaluate the MSP Program's role in student achievement, with two companion analyses. Although our pilot study uses a comparative approach, the study by Dimitrov in this issue follows a within-group design. The third analysis by Yin and his associates in this issue covers the varied designs used by the MSPs themselves in their own evaluations.

In this pilot, we focus on a sample of participating schools in one MSP in one state. The nonparticipating schools were carefully matched with the program participating schools on eight demographic variables to form a comparison group. This article offers detailed documentation on how we operationalize two matching methods for comparative purpose. We conclude that carefully executed matching methods are promising for large scale comparative analysis on the effects of the MSP Program across different states.



The purpose of this pilot study is to propose a set of analytical steps for comparing schools that participate in the National Science Foundation's Math and Science Partnership (NSF-MSP) Program and their nonparticipating peers in the same state. This pilot is part of a larger effort to evaluate the MSP Program's role in student achievement, with two companion analyses. Although our pilot study uses a comparative approach, the article by Dimitrov follows a within-group design. The third analysis by Yin and his associates covers the varied designs used by the MSPs themselves in their own evaluations. The overall objective of this larger effort is to examine whether the MSP Program is associated with student academic performance. In moving toward a comprehensive analysis of the outcomes associated with the program, we use state standardized test scores as a measure of student performance because of their public accessibility and prominence as accountability indicators. Ultimately, any conclusions drawn about the relationship between the MSP Program and student achievement is based on the convergence of all three analyses.

The purpose of the analysis is to examine whether MSP participating schools compared to non-MSP schools are associated with different achievement trends. Because MSP activities primarily involve teacher training and professional development in multiple grade levels, we examine school-level achievement. We address the question, When schools in a state participate in the MSP Program, do their students perform better than they would have if they had not participated in the MSP Program?

In placing the MSP participating schools in a comparative context, this pilot study uses social science methodologies that account for many confounding conditions, thereby making as fair a comparison as possible. Throughout the analysis, student achievement has been measured in terms of performance on state-administered assessments in mathematics and science for specific grades in the sampled schools.

Central to addressing the issue of school performance is accounting for a school's previous level of achievement. A simple comparison of MSP participating to non-MSP schools in a given year does not tell us about the potential association with the MSP engagement because it does not account for how those MSP participating and non-MSP schools were performing before the program began. Our statistical models, therefore, account for the prior level of achievement. We also consider factors such as student poverty levels and student family background, due to their documented association with student achievement outcomes.

It should be noted that the pilot comparison is still preliminary. Even though we have made every attempt to match demographically and academically similar MSP and non-MSP schools, the nature of any MSP-like activities in the non-MSP entities is still unknown. Because the MSP Program was not organized to follow a "treatment" and "no treatment" design, many of the non-MSP participating

schools may very well be undertaking MSP-like activities, using other sources of funds. In fact, the MSP entities in our study are limited to those funded by NSF, and our analysis has not yet had an opportunity to remove from the non-MSP group those districts and schools that might have received funding from the U.S. Department of Education as part of a counterpart MSP Program supported by that agency.

Future analyses will attempt to further define the non-MSP group more precisely. To the extent that data are available, our next step will differentiate within the non-MSP group those districts and schools known to have some MSP-like activities. Nevertheless, even though such sorting has not yet been possible because of a lack of needed data, the present pilot analysis provides an opportunity for testing the pertinent statistical methods on an otherwise appropriate array of information. Additional caveats surrounding the analysis are stated throughout the rest of this article.

## ANALYTIC DESIGN IN SCHOOL MATCHING

The MSP Program can be seen as an investment toward building the capacity of partnering schools and districts to improve teaching and learning in math and science. This pilot study focuses on the relationship between the MSP Program and one set of outcome measures, namely, standardized test scores. The MSP Program has provided for the opportunity to expand the capacity of schools and districts by bringing resources and commitment from institutions of higher education to support math and/or science curricula, teacher professional development, and increases in the highly qualified teacher supply. Equally important, the program is designed to build sustainable relationships between these K-12 school systems and other key institutions, including business and industry, professional organizations, state education agencies, and others with a stake in educational improvement (NSF, 2006).

### Research Model Structure

As emphasized by King, Keohane, and Verba (1994), the goal of social science research is inference. In our study, we wish to make inferences about the relationship between the MSP Program and concurrent student achievement trends in math and science. Theoretically, we want to look at the performance of an MSP participating school and compare it to the counterfactual: How would the school have performed without MSP participation? We cannot observe the counterfactual directly, but we use statistical methods designed to estimate the differences associated with MSP participation.

Our analysis includes MSP participating schools and matched non-MSP schools within the state in which the individual MSP is located. We match on student background and socioeconomic status variables. Overlooking variables such as these is a problem known as “omitted variable bias” and can lead to incomplete conclusions about the marginal differences associated with the MSP Program. Our pilot includes a measure of previous school achievement. Including baseline measures of achievement is critical for understanding the incremental difference associated with the MSP Program. It is not sufficient to know how an MSP participating school is doing this year; we want to know how it is doing this year, relative to previous baseline and programmatic years. Finally, our methods attempt to specify the uncertainty surrounding our estimates. Determining statistical significance is important for understanding how strong any MSP and non-MSP differences might be.

### Applying Mahalanobis Distance Matching

To control for a number of demographic variables, we employ the Mahalanobis distance matching to define an appropriate comparison school group before analysis (Xing & Rosenbaum, 1993). We first calculate what the average MSP participating school looks like over a set of eight variables and then use the Mahalanobis distance function to locate a group of like schools within the particular state.

The estimated statistical distance between the two  $N$  dimensional points is scaled by the statistical variation in each component of the point. For example, if  $\vec{x}$  and  $\vec{y}$  are two points from the same distribution which has covariance matrix,  $C$ , then the Mahalanobis distance is given by  $((\vec{x} - \vec{y})'C^{-1}(\vec{x} - \vec{y}))^{\frac{1}{2}}$  (Takeshita, Nozawa, & Kimura, 1993).

The resulting group of statistically “close” non-MSP schools is used as our comparison group for regression analysis. Though Mahalanobis distance matching is widely used in computer and spectrometry science, it is only beginning to be used in education policy studies (Good, Burross, & McCaslin, 2005).

### Alternative Types of Comparison Group Analysis

For this pilot study, we utilize Mahalanobis matching on a group of non-MSP schools using the “average” MSP school characteristics as the control variables to match on. We believe that this process provides us with the clearest general picture of significant effects attributed to the MSP Program because it allows us to match schools to a group of like schools while still providing a large enough sample to regress on. Our theory is that if we find significant effects of MSP involvement, we can double-check to ensure that other external conditions are not causing any

portion of the observed effect. Another reason to aggregate up to the “average” MSP school is to blunt possible outlying MSP school observations which are highly abnormal.

The following are alternative methodologies of creating non-MSP comparison groups:

1. Using Ordinary Least Squares (OLS) regression analysis to regress an entire universe of non-MSP schools against the MSP schools on multiple control variables. The strengths of this method are that it is commonplace in education research literature where large samples are attainable and it is comparatively effortless to complete once the model is created.

A significant problem with employing this technique is that it does not allow for the effects of grouping, that is, the problem of independence of observations. Because schools within a state/region tend to share certain characteristics (demographic, environmental, experiential, etc.), observations based on these schools are not fully independent. If we cannot ensure independence of observations, we cannot justify the use of OLS regression analysis. Second, this method does not produce a set of schools from which we can double-check for unquantifiable variables that may offset or mimic the effects of MSP involvement. Third, there is a possible issue with heteroskedasticity, that is, the error term could vary or increase with each observation. This being the case, the use of OLS to regress the entire universe of non-MSP schools against MSP schools is, yet again, invalidated (Becker & Hurn, 2004).

Because of these limitations of standard OLS, our use of matching is justified as a suitable quasi-experimental approach. Its strength is in the fact that it “includes a pretest as a covariate or matching variable [which] is better than an approach that does not.” The pretests reduce the bias (Boruch, 2007).

2. We could randomly select a subset of non-MSP schools for analysis. This would provide us with a set of schools from which we can double-check for unquantifiable variables that may offset or mimic the effects of MSP involvement, but it would not resolve the problems surrounding independence of observations and heteroskedasticity.
3. Third, we could use Mahalanobis distance matching to isolate a sister non-MSP school for every MSP school. We refer to this methodology as “one-to-one” matching later on in the article. This allows us to theoretically resolve the independence of observations and heteroskedasticity problems, but we can no longer employ an OLS regression model because there are only two observations. This being the case, any observed effect would be from net changes in student achievement from one year to the next between the two schools.



## OLS Regression

Once the distance matching score is computed, the pilot study employs OLS regression. OLS regressions have, for many years, been the standard statistical technique for evaluation in the field of educational policy (Hanushek, 1979). In keeping with economics literature surrounding the relationship between educational inputs and outputs, we assume an education production function (Hanushek, 1986). In this model, the outputs of school math and science achievement are seen as the function of a series of inputs. One of the inputs that some schools have, and others do not, is MSP participation. Our goal is to see if MSP participation is related to the outputs of math and science student achievement. The general form of the relationship is specified as

$$O_{it} = f(M_{it}) \quad (1)$$

where outcomes ( $O_{it}$ ) in school  $i$  in year  $t$  are understood to be a function of the vector of MSP Program activity ( $M_{it}$ ). We assume a linear form of the production function (Hanushek & Raymond, 2004; Hanushek, Rivkin, & Taylor, 1996).

Our linear estimation initially takes the following form:

$$O_{it} = \beta_o + \beta M_{it} + \varepsilon_{it} \quad (2)$$

## Considering the MSP Participation's Scope and Intensity

Thus far, we have only referred generally to a school's participation in an MSP. In our quantitative analysis, it is necessary to construct measures of MSP Program participation. We considered both the scope and intensity of the MSP participation. First, we used data from the MSP Management Information System (MSP-MIS) to identify the scope of each MSP. By scope, we mean both the subject (math and/or science) as well as the grade levels targeted. In many cases, an MSP's scope does not align precisely with the grades and subjects tested by state assessments.

In addition to scope, we also looked at the intensity of the MSP participation. It should be noted that this pilot study focused on a small set of MSP-MIS data to define the notion of intensity (Dimitrov, 2008/this issue). This data identified whether schools had met one of the following three conditions during school years 2002–03, 2003–04, or 2004–05:

- MSP-MIS item “q5Bald”: Whether 30% or more of targeted teachers participated in 30 or more hr of MSP-sponsored activities.

- MSP-MIS item “q5Bbld”: Whether 30% or more of targeted students engaged in a challenging mathematics or science curriculum that was initiated or revised with MSP support.
- MSP-MIS item “q5Bdld”: Whether 30% or more of targeted students participated in an MSP-supported academic enrichment activity.

For the purposes of this pilot study, if at least one of the three conditions was met for any given year, the school was considered as a “participating” MSP partner. Based on these considerations of scope and intensity, for each student achievement outcome measure available, we identified MSP participating schools as those schools that (a) were associated with an MSP activity targeted on the same grade-subject being tested by the state and (b) satisfied at least one of the three 30%-participating conditions.

### Variables for Statistical Matching

In addition to MSP participation, other school-level conditions are likely to be associated with student achievement. To address alternative explanations, we use relevant and available school-level control variables, as provided by state departments of education and the U.S. Department of Education’s National Center for Education Statistics’ Common Core of Data (NCES CCD).

Our first control variable is the size of the school, measured as the average total enrollment found in the NCES CCD data over the 3-year period. Larger schools operate under different conditions than smaller schools, in turn potentially influencing student achievement outcomes. Use of “size” as a control variable reduces, if not eliminates, any contaminating effect.

Second, the makeup of the school’s student body is likely connected to student performance. Schools/campuses serving larger percentages of Black and/or Latino students may experience lower overall achievement as they address the racial disparity that pervades the American public education (Jencks & Phillips, 1998). Another important control is for the percentage of students with disabilities in the school. Larger percentages of students with disabilities may be expected to reduce overall school achievement, as those students may face additional educational challenges.

We also include a measure of the percentage of students in the school/campus who are eligible for free and reduced-price lunch and Title I eligibility. Since the Coleman Report in 1965, a consistent finding in the social science literature on education is a strong relationship between family background and student success. The percentage of free/reduced-price lunch eligible students serves as a proxy for the students’ family background.

Third, the number of pupils per teacher in a given school was also taken into account. Education research exists that suggests that class-size reduction can benefit certain populations of students (Rivkin, Hanushek, & Kain, 2005).

Finally, we match on the locale of the school. We believe that the size and classification of the municipality where a school is located has a strong effect on how a school system operates and is structured. This can be directly linked to student achievement. Also, we posit that the type of neighborhood setting where students live has an influence on student achievement. The NCES CCD's categorical *Locale Code* variable was used to match non-MSP schools to the average MSP participating school. The NCES CCD (n.d.) glossary defines *Locale* as "the situation of a school in a particular location relative to populous areas, based on the school's address." The possible categories are Large City, Mid-Size City, Urban Fringe of Large City, Urban Fringe of Mid-Size City, Large Town, Small Town, Rural, outside [Core Based Statistical Area], and Rural, inside [Core Based Statistical Area]."

This pilot study decided not to include student mobility rates as a control variable because of the missing data from many schools in the pilot study state. As an aside, within those MSP participating and non-MSP schools from which we do have mobility data, the percentages are generally small, so it is improbable that it would have a significant effect on student outcomes. It is our intention to consider using this variable in further analyses.

## Measuring Achievement Gains

This pilot study assumes that a connection exists between increasing capacity and improvement in student performance. We measure student performance by examining state standardized test scores because these test scores are publicly accessible and allow us to collect multiple points of data over time to monitor trends regarding different schools. Both the direction and the magnitude of student achievement in specific subject areas and by grade levels can be informed by our preliminary analysis. It should be noted that school demographic characteristics and state standardized test performance are sensitive to the date of access to the Web sites at federal and state agencies. Data accuracy is dependent on the State Department of Education and U.S. Department of Education's NCES CCD.

In conducting our analysis, we measure the achievement gains from (or value-added by) MSP participation. In the literature that examines the effects of school funding on achievement, this is typically accomplished by modeling, either by generating a dependent variable that measures "change in performance from year  $t-1$  to year  $t$ " or by using performance in year  $t-1$  as a statistical control variable on the right-hand side of the equation (Burtless, 1996). We adopt the second approach, including lagged achievement as a statistical control variable. This

lagged achievement variable captures the average MSP participating school's performance in the previous year, relative to the matched non-MSP schools. One of the reasons we do not calculate a direct change-in-performance variable is that the test instrument in states may have changed over the time period of interest.

Introducing this notion of value-added through the use of a lagged achievement control variable enables us to better estimate the trends associated with the MSP Program, distinct from the influences of the eight variables chosen to match on, such as parental commitment to education. For instance, the assumption holds that if parental involvement is roughly the same year to year (e.g., active parents in year  $t-1$  are still active in year  $t$  and vice versa), then those parental involvement factors will be captured by the lagged achievement variable. However, if parent involvement also changes from year to year (but such data are not available) and systematically with achievement, adjusting for such a contamination would be outside of our model's capability. Overall, given the available data, we believe this is the most complete model we can develop.

### School-Level OLS Model

In our OLS regressions, we employed STATA's *rreg* command to obtain robust regression estimates.<sup>1</sup> Our school-level statistical OLS regression model takes the form

$$\text{ACHIEVE}_{jt} = \beta_0 + \beta_1 \text{ACHIEVE}_{jt-1} + \beta_2 \text{MSP}_{jt} + \varepsilon_{jt} \quad (3)$$

where  $\text{ACHIEVE}_{jt}$  is the math or science student achievement score for school  $j$  in year  $t$ ,  $\text{ACHIEVE}_{jt-1}$  is the school's previous achievement level,  $\text{MSP}_{jt}$  is a dichotomous (dummy) variable indicating whether this is an MSP participating school (after accounting for MSP participation scope and intensity; discussed earlier), and  $\varepsilon_{jt}$  is the error term. We use this base model for our school-level analysis and can apply this to all grade levels where we have such data.

### MATCHING OF COMPARISON SCHOOLS WITH MSP PARTICIPATING SCHOOLS

We piloted this study in a state where there is broad access to school-level data. The state also has multiple operating MSPs (three in this case). One of the three MSPs located within the state is a Cohort II awardee, so it was dropped from the

---

<sup>1</sup>This is a log-rank weighting function. For further explanation as to *rreg*'s methodology, see *Statistics With STATA* by Hamilton (2006).



TABLE 1  
Math and Science Partnership Participating Schools in One State

School	Year 1			Year 2			Year 3		
	Question 1	Question 2	Question 3	Question 1	Question 2	Question 3	Question 1	Question 2	Question 3
Middle School 1	Yes	No	No	Yes	No	No	Yes	No	No
Middle School 2	Yes	No	No	Yes	No	No	Yes	No	No
Middle School 3	No	No	No	Yes	No	No	Yes	No	No
Middle School 4	No	No	No	No	No	No	Yes	No	No
Middle School 5	No	No	No	Yes	No	No	Yes	No	No
Middle School 6	Yes	No	No	Yes	No	No	Yes	No	No
Middle School 7	No	No	No	Yes	No	No	Yes	No	No
Middle School 8	Yes	No	No	No	No	No	Yes	No	No
Middle School 9	Yes	No	No	Yes	No	No	Yes	No	No
Middle School 10	No	No	No	Yes	No	No	Yes	No	No
Middle School 11	Yes	No	No	Yes	No	No	Yes	No	No
Middle School 12	No	No	No	Yes	No	No	Yes	No	No
High School 1	No	No	No	Yes	No	No	Yes	No	No
High School 2	Yes	No	No	Yes	No	No	Yes	No	No
High School 3	No	No	No	No	No	No	Yes	No	No
High School 4	No	No	No	Yes	No	No	Yes	No	No
High School 5	No	No	No	No	No	No	Yes	No	No
High School 6	Yes	No	No	Yes	No	No	Yes	No	No
High School 7	Yes	No	No	No	No	No	Yes	No	No
High School 8	Yes	No	No	Yes	No	No	Yes	No	No
High School 9	Yes	No	No	Yes	No	No	Yes	No	No
High School 10	Yes	No	No	Yes	No	No	Yes	No	No
High School 11	No	No	No	Yes	No	No	Yes	No	No

Note. Answered yes to Management Information System questions.

TABLE 2  
Matching Variables

Variable	Source	Abbreviation
Total enrollment	NCES CCD	ENROLL
% Black/African American students	NCES CCD <sup>a</sup>	PCT_BLK
% Latino students	NCES CCD <sup>a</sup>	PCT_LATN
% students with disabilities	State data	PCT_SWD
School Title I eligibility	NCES CCD	TITLE1
% students free/reduced lunch eligible	NCES CCD	PCT_LUNCH
Students-to-teacher ratio	NCES CCD	PPT
Locale of school	NCES CCD	LOCALE
School mobility rate <sup>b</sup>	State data	MOBILITY

<sup>a</sup>Percentage calculated from a raw number. <sup>b</sup>Variable dropped due to extensive missing data.

analysis because it is desirable to have at least three years of student achievement data to analyze: one year as a baseline and two subsequent project years.

According to the MSP-MIS data, and using the definition of “intensity” of MSP participation as previously discussed, one of the two remaining MSPs did not have any participating schools and the other had 24 schools (out of 39 schools) participating. One of the 24 schools is not defined as a “regular school” by the NCES CCD and does not report student achievement data as the others do. This being the case, it was dropped from our analysis. Of the remaining schools, 12 are considered middle schools and 11 are high schools.<sup>2</sup> Table 1 displays the results of the “intensity” analysis.

As previously mentioned, the characteristics in Table 2 were used to match MSP participating schools to non-MSP (comparison) schools. These variable data were taken from the MSP participating schools and aggregated-up to construct an average MSP school profile to match a set of non-MSP schools. The variable profile of a particular school was constructed from averaging the variables for all three school years, 2002–03, 2003–04 and 2004–05 (see National Center for Education Statistics, Common Core of Data, <http://nces.ed.gov/ccd>). The descriptive variable characteristics of the MSP participating schools are as in Tables 3 and 4. Tables 5 and 6 are placed after Tables 3 and 4 for comparative purposes.

Once we construct the average MSP participating middle and high school profiles, we match the entire universe of the state’s traditional public non-MSP middle and high schools using Mahalanobis distance matching. This methodology calculates a distance for each non-MSP school signifying how well it matches the

<sup>2</sup>The particular MSP analyzed was targeted to focus only on middle and high schools, not elementary schools.

TABLE 3  
Descriptive Statistics for Math and Science Partnership Middle Schools

Variable	$M(\mu)$	$SD(\sigma)$	Min	Max
ENROLL	678	290	320	1401
PCT.BLK	5.42%	5.24%	0.12%	14.63%
PCT.LATN	0.45%	0.19%	0.12%	0.89%
PCT.SWD	13.77%	4.19%	6.97%	20.53%
TITLE1 <sup>a</sup>	0.67	0.49	0.00	1.00
PCT.LUNCH	19.95%	15.32%	0.00%	46.48%
PPT	17	2	14	21
LOCALE <sup>b</sup>	4	2	2	8

Note.  $N = 12$ . <sup>a</sup>Dummy variable where 1 = Title I Eligible School. <sup>b</sup>1 = Large City, 2 = Mid-Size City, 3 = Urban Fringe of Large City, 4 = Urban Fringe of Mid-Size City, 5 = Large Town, 6 = Small Town, 7 = Rural, outside [Core Based Statistical Area], 8 = Rural, inside [Core Based Statistical Area].

average middle and high school MSP school profile. In short, it determines the range between the minimum and maximum values for each variable and then sums-up all of those values. That number becomes the maximum distance value. So, a perfect match would have a distance value of 0 and the furthest observation from the average MSP school profile would have a distance value equal to the maximum distance value.

Tables 7 and 8 display the top 10 matches from the entire universes of the state's traditional public non-MSP middle and high schools. The middle school described

TABLE 4  
Descriptive Statistics for Math and Science Partnership High Schools

Variable	$M(\mu)$	$SD(\sigma)$	Min	Max
ENROLL	1071	427	644	2039
PCT.BLK	2.19%	0.32%	0.00%	8.98%
PCT.LATN	0.53%	0.28%	0.05%	0.92%
PCT.SWD	12.65%	3.38%	6.57%	17.00%
TITLE1 <sup>a</sup>	0.09	0.30	0.00	1.00
PCT.LUNCH	14.96%	10.97%	5.51%	46.40%
PPT	17	2	14	20
LOCALE <sup>b</sup>	5	2	2	8

Note.  $N = 11$ . <sup>a</sup>Dummy variable where 1 = Title I Eligible School. <sup>b</sup>1 = Large City, 2 = Mid-Size City, 3 = Urban Fringe of Large City, 4 = Urban Fringe of Mid-Size City, 5 = Large Town, 6 = Small Town, 7 = Rural, outside [Core Based Statistical Area], 8 = Rural, inside [Core Based Statistical Area].

TABLE 5  
Descriptive Statistics for Non-Math and Science Partnership Middle Schools

Variable	<i>M</i> ( $\mu$ )	<i>SD</i> ( $\sigma$ )	Min	Max
ENROLL	697	308, 345	1498	
PCT_BLK	6.08%	0.68%	0.29%	18.62%
PCT_LATN	0.83%	0.40%	.20%	1.89%
PCT_SWD	14.29%	3.86%	7.23%	21.17%
TITLE1 <sup>a</sup>	0.67	0.49	0.00	1.00
PCT_LUNCH	20.86%	14.58%	0.00%	51.03%
PPT	16	2	14	19
LOCALE <sup>b</sup>	4	2	2	8

*Note.* *N* = 10. <sup>a</sup>Dummy variable where 1 = Title I Eligible School. <sup>b</sup>1 = Large City, 2 = Mid-Size City, 3 = Urban Fringe of Large City, 4 = Urban Fringe of Mid-Size City, 5 = Large Town, 6 = Small Town, 7 = Rural, outside [Core Based Statistical Area], 8 = Rural, inside [Core Based Statistical Area].

TABLE 6  
Descriptive Statistics for Non-Math and Science Partnership High Schools

Variable	<i>M</i> ( $\mu$ )	<i>SD</i> ( $\sigma$ )	Min	Max
ENROLL	998	75	886	1119
PCT_BLK	3.25%	.39%	0.42%	13.38%
PCT_LATN	0.62%	0.20%	0.20%	1.00%
PCT_SWD	14.48%	4.67%	6.76%	23.23%
TITLE1*	0	0	0	0
PCT_LUNCH	10.00%	6.13%	2.76%	21.52%
PPT	17	1	15	20
LOCALE**	5	1	4	6

*Note.* *N* = 10. <sup>a</sup>Dummy variable where 1 = Title I Eligible School. <sup>b</sup>1 = Large City, 2 = Mid-Size City, 3 = Urban Fringe of Large City, 4 = Urban Fringe of Mid-Size City, 5 = Large Town, 6 = Small Town, 7 = Rural, outside [Core Based Statistical Area], 8 = Rural, inside [Core Based Statistical Area].

in Table 9 is an example of a not close match to the average MSP participating middle school profile. Tables 5 and 6 display the descriptive statistics for these 10 non-MSP middle schools and 10 non-MSP high schools.

ALTERNATIVE ONE-TO-ONE MATCHING METHODOLOGY

Instead of matching a set number of non-MSP schools to the “average” characteristics of the MSP participating school sample, we can also match each MSP



TABLE 7  
Variable Profiles of Matched Non-Math and Science Partnership Middle Schools

School	ENROLL	PCT_BLK	PCT_LATN	PCT_SWD	TITLE1 <sup>a</sup>	PCT_LUNCH	PPT	LOCALE <sup>b</sup>	DISTANCE <sup>c</sup>
Average MSP MS	618	5.23%	0.46%	11.68%	0.58	22.27%	17	5	0.0000
Matched MS 1	629	0.00%	1.55%	13.37%	1.00	24.38%	16	4	0.9332
Matched MS 2	556	1.44%	0.31%	10.80%	1.00	24.92%	13	4	0.9951
Matched MS 3	493	0.91%	0.20%	13.20%	1.00	20.39%	18	4	1.0134
Matched MS 4	618	0.57%	0.45%	8.17%	1.00	12.91%	17	4	1.0234
Matched MS 5	602	6.39%	0.37%	10.97%	1.00	11.83%	17	4	1.0280
Matched MS 6	691	2.93%	0.33%	11.83%	1.00	11.51%	15	4	1.0511
Matched MS 7	549	4.96%	0.68%	13.90%	1.00	14.30%	14	4	1.0706
Matched MS 8	479	0.68%	0.00%	10.83%	1.00	16.91%	17	4	1.0807
Matched MS 9	589	5.60%	1.10%	9.43%	1.00	8.27%	17	4	1.0874
Matched MS 10	608	14.10%	1.03%	12.90%	1.00	32.80%	14	4	1.1159

<sup>a</sup>Dummy variable where 1 = Title I Eligible School. <sup>b</sup>1 = Large City, 2 = Mid-Size City, 3 = Urban Fringe of Large City, 4 = Urban Fringe of Mid-Size City, 5 = Large Town, 6 = Small Town, 7 = Rural, outside [Core Based Statistical Area], 8 = Rural, inside [Core Based Statistical Area].

<sup>c</sup>Out of a possible 361.

TABLE 8  
Variable Profiles of Matched Non-Math and Science Partnership High Schools

School	ENROLL	PCT_BLK	PCT_LATN	PCT_SWD	TITLE1 <sup>a</sup>	PCT_LUNCH	PPT	LOCALE <sup>b</sup>	DISTANCE <sup>c</sup>
Matched HS 1	954	1.10%	0.47%	11.27%	0	21.52%	16	4	0.4217
Matched HS 2	947	2.61%	0.61%	11.97%	0	5.60%	19	4	0.4598
Matched HS 3	1,065	4.44%	0.56%	15.97%	0	15.36%	19	6	0.4623
Matched HS 4	992	4.74%	0.66%	12.03%	0	12.00%	20	6	0.4646
Matched HS 5	1,060	0.54%	0.71%	9.03%	0	2.76%	18	4	0.4755
Matched HS 6	885	2.01%	0.20%	13.53%	0	13.33%	17	4	0.4836
Matched HS 7	1,043	1.13%	0.72%	13.70%	0	5.54%	17	6	0.4929
Matched HS 8	908	2.12%	0.72%	8.37%	0	4.65%	17	4	0.5194
Matched HS 9	1,007	13.38%	0.97%	8.77%	0	4.92%	17	4	0.5195
Matched HS 10	1,119	0.42%	0.63%	11.30%	0	13.99%	15	6	0.5209

<sup>a</sup>Dummy variable where 1 = Title I Eligible School. <sup>b</sup>1 = Large City, 2 = Mid-Size City, 3 = Urban Fringe of Large City, 4 = Urban Fringe of Mid-Size City, 5 = Large Town, 6 = Small Town, 7 = Rural, outside [Core Based Statistical Area], 8 = Rural, inside [Core Based Statistical Area]. <sup>c</sup>Out of a possible 368.

TABLE 9  
Variable Profile of Non-Matched Non-Math and Science Partnership Middle School

School	ENROLL	PCT_BLK	PCT_LATN	PCT_SWD	TITLE1 <sup>a</sup>	PCT_LUNCH	PPT	LOCALE <sup>b</sup>	DISTANCE <sup>c</sup>
Average MSP MS	618	5.23%	0.46%	11.68%	0.58	22.27%	17	5	0
Nonmatched MS	809	17.32%	62.12%	21.27%	1.00	93.72%	11	1	200.8766

<sup>a</sup>Dummy variable where 1 = Title I Eligible School. <sup>b</sup>1 = Large City, 2 = Mid-Size City, 3 = Urban Fringe of Large City, 4 = Urban Fringe of Mid-Size City, 5 = Large Town, 6 = Small Town, 7 = Rural, outside [Core Based Statistical Area], 8 = Rural, inside [Core Based Statistical Area].

<sup>c</sup>Out of a possible 361.

TABLE 10  
Variable Profiles of One-to-One Matched Middle School Pairs

Pair	School	ENROLL	PCT_BLK	PCT_LATN	PCT_SWD	TITLE1 <sup>a</sup>	PCT_LUNCH	PPT	LOCALE <sup>b</sup>	DISTANCE
Pair 1	MSP	397	1.32%	0.19%	11.76%	1	0.00%	21	4	0.0000
	Match	455	0.38%	0.49%	9.00%	1	11.15%	19	4	0.2709
Pair 2	MSP	794	3.21%	0.44%	15.33%	0	19.85%	18	4	0.0000
	Match	866	3.69%	0.63%	16.66%	0	24.62%	14	4	0.2069
Pair 3	MSP	625	8.20%	0.44%	16.70%	1	33.48%	17	4	0.0000
	Match	608	14.10%	1.03%	12.06%	1	32.80%	14	4	0.0895
Pair 4	MSP	320	0.16%	0.47%	13.33%	0	0.00%	17	8	0.0000
	Match	359	0.35%	0.77%	14.00%	0	0.00%	15	8	0.0608
Pair 5	MSP	1,411	2.52%	0.43%	6.97%	1	7.44%	18	4	0.0000
	Match	1,498	9.21%	1.89%	13.40%	1	11.21%	18	3	0.4681
Pair 6	MSP	813	0.22%	0.58%	11.10%	1	6.15%	17	8	0.0000
	Match	819	1.68%	0.92%	16.66%	1	13.16%	17	8	0.2383
Pair 7	MSP	545	13.22%	0.55%	20.53%	1	46.49%	16	2	0.0000
	Match	508	16.63%	0.20%	21.16%	1	51.03%	14	2	0.0681
Pair 8	MSP	809	0.12%	0.12%	10.16%	1	23.19%	18	4	0.0000
	Match	807	5.23%	0.84%	14.46%	1	17.50%	17	4	0.1102
Pair 9	MSP	687	11.07%	0.51%	16.43%	0	25.02%	18	4	0.0000
	Match	721	1.94%	0.87%	12.43%	0	18.13%	16	4	0.1998
Pair 10	MSP	777	3.77%	0.42%	17.33%	1	26.63%	18	2	0.0000
	Match	754	0.80%	0.63%	16.30%	1	25.73%	15	2	0.1557
Pair 11	MSP	369	6.58%	0.34%	8.10%	0	11.32%	15	4	0.0000
	Match	345	0.29%	0.80%	7.23%	0	6.01%	14	3	0.3204
Pair 12	MSP	593	14.63%	0.89%	17.43%	1	39.84%	14	4	0.0000
	Match	626	18.62%	0.92%	18.06%	1	38.95%	14	4	0.0306

*Note.* MSP = Math and Science Partnership. Dummy variable where 1 = Title I Eligible School. <sup>b</sup>1 = Large City, 2 = Mid-Size City, 3 = Urban Fringe of Large City, 4 = Urban Fringe of Mid-Size City, 5 = Large Town, 6 = Small Town, 7 = Rural, outside [Core Based Statistical Area], 8 = Rural, inside [Core Based Statistical Area].



participating school to a *similar* school outside of the MSP program (Rubin & Thomas, 1996). The same Mahalanobis distance matching methodology can be utilized to produce this match or “sister school” for each of our MSP participating schools.

Next are middle school examples of the results from the school-to-school matching pilot study. Each school pair is matched on the same eight variables as listed in Table 2. In using a one-to-one matching protocol (i.e., one school vs. another), we can directly compare standardized test scores from one year to the next. Because there are only two observations, we cannot regress using OLS. Table 10 displays the middle school pairs’ test scores and reports if there is any difference between them.

## IMPLICATIONS AND FUTURE WORK

In this pilot, we focus on a sample of MSP participating schools in one MSP located in one state. The non-MSP participating schools were carefully matched with the program participating schools on eight demographic variables to form a comparison group. This article offers detailed documentation on how we operationalize two matching methods for comparative purpose. This is compliant with the U.S. Department of Education’s Academic Competitive Council’s charge to evaluate the effectiveness of science, technology, engineering, and mathematics education intervention under rigorous conditions. In a hierarchy with “Experimental Methods such as Randomized Controlled Trials” at the top and “Other designs, such as Pre- and Post-Test Studies, and Comparison Group Studies without careful matching” at the bottom, our matching methodology falls in between as one that is a “Quasi-experimental Method such as Well-Matched Comparison Group Study” (U.S. Department of Education, 2007). In the absence of having the prime condition of being able to conduct a randomized controlled trial of MSP-funded schools, we will continue to refine our matching methodology to provide the most appropriate quasi-experimental method so that it may act as a model for similar program analyses.

We believe that when an MSP is implemented in certain types of schools, the variation of variables across observations is small, so matching up to the aggregate average MSP school is an efficient matching method. Of importance, this allows for regression analysis, though our sample sizes will likely be small. Conversely, if the variation across observations is large, it would be more prudent to turn to the one-to-one matching methodology. Although the method will no longer allow for regression analysis, it will provide us with the nearest non-MSP neighboring school for direct comparison on student achievement output effects. As we pilot this matching protocol across the Cohort I MSP schools, we will utilize both matching methods and compare and contrast the results. This will better inform our hypothesis of which matching technique is best suited for our further analyses.

Our matching results suggest that carefully executed matching methods are promising for large-scale comparative analysis on the effects of MSP programs across different states. Our next step is to expand the methods to include other states and additional data. Ultimately, the goal is to analyze the relationship between MSP school participation and state standardized achievement test gains. We shall do this through our matching methodology that controls for various extraneous factors that may affect student test scores. We will employ both standard and robust OLS regression analysis on the sets of middle and high schools. The analyzed data will also be from school years 2002–03, 2003–04, and 2004–05. Further, grade-level state standardized test scores in both math and science in school year 2002–03 will be used as the baseline, as this was Year 1 of MSP award funding.

Our research team also plans to continue to expand the list of control variables beyond the eight that we have used in the pilot. Among the additional control variables of interest at the school level are federal grants in math and science, teacher and principal turnover rates, and other organizational conditions. Our effort in defining and operationalizing an appropriate comparison group in the MSP evaluation study will contribute to a broader discussion in program evaluation.

## ACKNOWLEDGMENT

This pilot study is one in a series of substudies for the Math and Science Partnership Program Evaluation (MSP-PE) conducted for the National Science Foundation's MSP Program. The MSP-PE is conducted under Contract No. EHR-0456995. Since 2007, Bernice Anderson, Ed.D., Senior Advisor for Evaluation, Directorate for Education and Human Resources, has served as the National Science Foundation Program Officer. The authors are Kenneth K. Wong, Ph.D., of Brown University, and Ted Socha, M.A., of National Center for Education Statistics.

This illustrative study draws on publicly accessible school-level data files from one state and from data available at the National Center for Education Statistics' Common Core of Data. In addition, the study consults secondary materials only: available literature together with all of the annual reports, evaluation reports, MSPnet documents, and Web site information reported by the individual Math and Science Partnerships (MSPs) in the MSP Program accessible through the school year 2004–05.

We express gratitude to the Brown research team consisting of Joshua Marland and Erikson Arcaira. Francis X. Shen also provided valuable support on data analysis.

## REFERENCES

- Becker, R., & Hurn, A. S. (2004). Testing for nonlinearity in mean in the presence of heteroskedasticity. School of Economics and Finance. Queensland University of Technology.

- Boruch, R. (2007). Encouraging the flight of error: Ethical standards, evidence standards, and randomized trials. In G. Julnes & D. J. Rog (Eds.), *Informing federal policies on evaluation methodology: Building the evidence base for method choice in government sponsored evaluation* (no. 113, pp. 55–73). San Francisco: Jossey-Bass.
- Burtless, G. (Ed.). (1996). *Does money matter? The effect of school resources on student achievement and adult success*. Washington, DC: Brookings.
- Dimitrov, D. M. (2008/this issue). Initial trends in MSP-related changes in student achievement with MIS data. *Peabody Journal of Education*, 83, 637–653.
- Good, T. L., Burross, H. L., & McCaslin, M. M., (2005). Comprehensive school reform: A longitudinal study of school improvement in one state. *Teachers College Record*, 107(10), 2205–2226.
- Hamilton, L. C. (2006). *Statistics with STATA* (6th ed.). Pacific Grove, CA: Duxbury Press.
- Hanushek, E. A. (1979, Summer). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources*, pp. 351–388.
- Hanushek, E. A. (1986, September). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, pp. 1141–1177.
- Hanushek, E. A., & Raymond, M. F. (2004). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24(1), 297–328.
- Hanushek, E. A., Rivkin, S. G., & Taylor, L. L. (1996). Aggregation and the estimated effects of school resources. *The Review of Economics and Statistics*, 78(4), 611–627.
- Jencks, C., & Phillips, M. (Eds.). (1998). *The Black-White test score gap*. Washington, DC: Brookings Institution Press.
- King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton, NJ: Princeton University Press.
- National Center for Education Statistics. (n.d.). *CCD school and district glossary*. Washington, DC: Author. Retrieved from <http://nces.ed.gov/ccd/commonfiles/glossary.asp>
- National Science Foundation. (2006). *Math and Science Partnership Program* (NSF-05-069). Arlington, VA: Author. Retrieved May 8, 2007, from <http://nsf.gov/pubs/2005/nsf05069/nsf05069.pdf>
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458.
- Rubin, D. B., & Thomas, N. (1996). Matching using estimated propensity scores: Relating theory to practice. *Biometrics*, 52(1), 249–264.
- Takeshita, T., Nozawa, S., & Kimura, F. (1993). *On the bias of mahalanobis distance due to limited sample size effect*. Paper presented at the Second Conference on the Institute of Electrical and Electronics Engineers, Tsukuba Science City, Japan.
- U.S. Department of Education. (2007). *Report of the Academic Competitiveness Council*. Washington, DC: Author.
- Xing, S. G., & Rosenbaum, P. R. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. *Journal of Computational and Graphical Statistics*, 2(4), 405.

# Discovering “What’s Innovative”: The Challenge of Evaluating Education Research and Development Efforts

Robert K. Yin  
*COSMOS Corporation*

Edward J. Hackett  
*Arizona State University*

Daryl E. Chubin  
*American Association for the Advancement of Science*

## THE MATH AND SCIENCE PARTNERSHIP (MSP) PROGRAM’S DUAL NICHE

National Science Foundation’s (NSF’s) MSP Program seeks foremost “to improve student outcomes in high-quality mathematics and science by all students, at all pre-K-12 levels” (NSF-02-061; NSF, 2001). In fact, the program aims to impact large numbers of such students. At the same time, the MSP Program positions itself as “a major research and development effort” (NSF-03-605; NSF, 2003). Presumably, as a research and development (R&D) effort, the MSP Program should produce new ideas or innovations that will eventually impact how students learn.

These two conditions define a dual niche for the MSP Program. The program is concerned with implementing change in existing pre-K-12 systems as well as discovering new (and improved) ways of providing pre-K-12 education. The dual niche poses a challenge to evaluators and the design of evaluations



such as the MSP Program Evaluation that is the subject of this entire special issue.

The need is to develop ways of assessing how well the R&D function is working, not just whether existing pre-K-12 systems have been impacted. The evaluation challenge is complicated by the protracted and potentially lengthy character of R&D as a process, which also often reveals only glimmers of promise at the outset and unforeseen pathways thereafter.

The challenge complements other strategies for evaluating education programs. In particular, during the past several years federally-supported K-12 education research and evaluation has veered toward identifying effective education practices and developing empirical evidence on those practices. "*What Works?*" has become a shorthand term for this quest, and the federal investments have included a "What Works" clearinghouse (<http://www.whatworks.ed.gov/>) of best practices and supportive research evidence to be shared with school systems across the country. In this endeavor, randomized control trials (RCTs) have been enshrined as the strongest research design for demonstrating the effectiveness of an educational intervention. To perform an RCT, investigators randomly assign the individuals or groups being studied to the various "treatment" and "control" conditions, emulating the successful use of these designs in clinical trials in medicine (Jadad, 1998). For these reasons RCTs are at the heart of "scientifically based" research designs in education (e.g., Shavelson & Towne, 2002; Towne & Hilton, 2004; Towne, Wise, & Winters, 2004).

Unfortunately, the pursuit of effective practices has become so pervasive that many people, including policymakers, may have begun to believe that the quest represents the totality of scientific education research and evaluation. Overshadowed by such a belief are the significant advances in education produced by the complementary—and in many ways more traditional—line of scientific education research aimed at discovering new ways to educate and learn. "*What's Innovative?*" might be shorthand for understanding an alternative approach to scientific education research.

The innovations might come from varied sources. They include R&D programs such as the MSP Program but also discoveries revealed in the course of education practice. To give but a sense of the potential breadth and heritage of such innovations, the appendix offers a list of "candidate" education innovations gleaned informally and not exhaustively from the education literature.

The list deliberately mingles conceptual and practical advances. The point is that all of them originated through some discovery or innovation processes. Whether and how such processes can be monitored and evaluated serves as the challenge for designing the MSP Program Evaluation to assess the R&D function of the MSP Program. Our study clarifies the kinds of advances that are being valued, then suggests a possible evaluation strategy.

## DISCOVERIES AND INNOVATIONS

### Discoveries and Innovations in Science

Taken literally, the word *discovery* implies removing the cover from something that already exists but was unknown, as one might discover a new species, element, mountain, tribe, and such. But a closer look at discovery reveals not a single act but a *process* of searching, making, justifying, challenging, validating, recognizing, rewarding, and building—an intensely social enterprise that has much in common with invention and other creative activities (Brannigan, 1981).

Perhaps this is why most discoveries are multiples, achieved simultaneously by researchers working independently. It is the flow of questions, speculations, investigations, and methods engaging a community of researchers that both gives rise to the claim and fuels the process of validation and acceptance (Merton, 1973). Finding an isolated stone-age tribe would seem to be a discovery, but its meaning and significance rest in what our study of the tribe teaches us about the development of the human species or the variety and sequence of its forms of social organization (or its use of tools, forms of family organization, etc.). Such claims rest on creative arguments and their acceptance by a community of scientists. In other words, the essence of discovery resides not only (or perhaps principally) in the thing uncovered but also in its implications for things already known, and in the meanings such things impart to the discovery. Such meanings arise from the community of scientists. This is why discovery, in its fullest sense, must be understood as a communal accomplishment.

Other discoveries are equally reliant on social process and intellectual context: sexual recombination in bacteria, transposons in corn, and “Lucy” and other human ancestors are all celebrated discoveries. Yet what we celebrate are the ideas, insights, explanations, and surprises these discoveries offer. Those in turn are not inherent but are created or invented using ideas, methods, and analytic techniques generally accepted in—that is, ratified by—a field. Making a discovery is a crucial and delicate social process that begins with a bold assertion and continues through the communal processes of evaluation, dissemination, evaluation, replication, accumulation, and incorporation or rejection (see Collins, 1985; Fleck, 1981; Latour & Woolgar, 1979). In much the same way, the meaning and significance of education discoveries reside not only in what is learned but also in what is done with the knowledge.

Some discoveries are resisted because they arrive prematurely (van Raan, 2004). As Stent (2002) pointed out, one way a discovery may be premature is if there is no theory available to explain it. In this case it simply seems implausible. Semmelweis’s ideas about how disease was transferred from cadavers to mothers who had just given birth were difficult to fathom in the absence of germs and germ theory. Mendel’s experiments were performed several decades before the idea

of genes was created, so the results lacked context, meaning, and consequences. Polanyi, Einstein, and others all proposed ideas that were ahead of their time.

In contrast, some discoveries are predictable or expected—the discovered phenomenon is known to exist—so when the supporting evidence passes muster the discovery rapidly wins acceptance. Examples would include the structure of the DNA molecule and the identification of trans-uranium elements. Each was an expected discovery in harmony with theory and research of its time, so once technical aspects of the evidence were judged satisfactory the discovery was appropriately welcomed. Notably, the practical implications and uses of such discoveries remain unsettled, and the ethical sensibilities to guide their development and application may only be in their formative stages (Bell, 1980).

Other discoveries are not known to exist but when found turn out to be consistent with accepted theories and their reasonable extensions. Genes associated with diseases (e.g., breast cancer) are now in this category: Theory does not indicate exactly what will be found, but the findings are consistent with current explanations and understandings about disease. Still other discoveries contradict accepted knowledge and meet with intense resistance (*heliobacter pylori*, the bacteria that cause ulcers, is a good example, as are the early cancer viruses and reverse transcriptase; on *heliobacter* see Thagard, 1999, chaps. 3–6).

Overall, accomplishments of the following sort may accompany discoveries and innovations in science (Clinedinst, 2005):

- Leading to a new line of inquiry within a field
- Solving a long-standing, important problem
- Filling a significant knowledge gap
- Advancing a theoretical framework guiding research in a field
- Developing a new research instrument or technique

## Sources of Discoveries and Innovations in Education

The criteria and processes just discussed have been drawn from historical and social studies of science. However, the sciences and education R&D may contribute to our understanding of educational practice more narrowly.

In recent years the range of sciences that conduct education-relevant research has expanded significantly. As Feuer (2004) observed,

the cognitive revolution in psychology has led to breakthroughs in our understanding of how human beings learn and to models for training and education (in schools as well as in business organizations and the military).

Other examples of rigorous research that have had an impact on public education are in the areas of measurement (testing and assessment), program evaluation, teaching of reading and mathematics, understanding the effects of race and class on educational

attainment, the effects of computer and other information technologies on academic achievement, and the economics of resource allocation in schools.

With the growing relevance of cognitive science, psychology, and neuroscience for understanding human learning and intelligence, it may be difficult today to place disciplinary bounds around education R&D. The Harvard Graduate School of Education, for example, offers a yearlong course titled Education, Psychology, and Neuroscience to help students understand the disciplinary variety of research and theory on human learning. This is a small outcropping of a larger process of “interdisciplinarity” (Klein, 1996). Not only is education R&D interdisciplinary, but it also stretches from the realm of fundamental knowledge into an emergent and heterogeneous knowledge of practice. In the words of one researcher,

Our science forces us to deal with particular problems, where local knowledge is needed. Therefore, ethnographic research is crucial, as are case studies, survey research, time series, design experiments, action research, and other means to collect reliable evidence for engaging in unfettered argument about education issues. (Berliner, 2002, p. 20)

This eclecticism poses novel opportunities for researchers and distinctive challenges for reviewers. The opportunities arise from the creative energy of working in “Pasteur’s quadrant,” that zone of use-inspired fundamental research that adds urgency and concreteness to the purposes of research, producing a distinctive form of originality (Stokes, 1997). But the challenges emerge from the way the quality of ideas and findings will be entangled with their uses and implications, an entanglement that may trigger political or value judgments.

Other disciplines in addition to education may generate discoveries of the types encountered in K-12 science, mathematics, engineering, and technology (STEM) research. Shavelson and Towne (2002, p. 6) acknowledged a pluralism of sources this way:

The design of a study does not make the study scientific. A wide variety of scientific designs are available for education research. They range from randomized experiments of voucher programs to in-depth ethnographic case studies of teachers to neurocognitive investigations of number learning using positive emission tomography brain imaging.

To us, “sources” and “designs” reside in different “education worlds.” Simply put, discoveries and innovations originate from *education R&D* (and evaluation), from *scientific research* in various fields (but especially mathematics, psychology, and the hybrids such as neuroscience), and from *classroom practice*. There are



different practitioners inhabiting these worlds, with only the latter developing insights as faculty (at all levels, K-16 and beyond). From these three different contexts come discoveries that more than pass review muster—they are hailed in the literature as salient (though their evidence bases vary a lot), if not significant for advancing, indeed changing, teaching and/or learning.

Possibly in recognition of this diversity of contributors, the MSP Program takes as its main tenet the engagement of STEM discipline faculty in mathematics and science education at all levels. Toward this end, the program has specified the “core partnership” for every funded MSP to be a collaboration between at least one local school district and one institution of higher education (IHE), and in particular the faculties in the STEM disciplines.

In a sense, education R&D may be “colonized” by researchers from cognate fields, and discoveries within the field of education may be catalyzed by developments elsewhere. Although the word *interdisciplinarity* describes cross-fertilization and collaboration across disciplinary and research lines, it does not indicate intellectual dominance, that is, which field is the colonizer and which the colonized. Education as the *colonizer*—as the field that borrows, imports, or annexes discoveries (in one of the multiple forms just described)—surveys a broad research terrain that not only may enrich the knowledge base for teaching and learning but also may change classroom education practice.

## COMMON PROCESSES

Regardless of academic field—or, for that matter, education versus non-education R&D—contemporary researchers follow common processes in pursuing discoveries and innovations. Any of four processes, pursued alone or in any combination, may be relevant and are described in Table 1: uncovering, inventing, explaining, and substantiating. Key to these processes is the constant interplay between claim and counterclaim that should permeate any discourse about these processes (Kelly & Yin, 2007). Thus, to assess the MSP Program as an R&D effort, one place to start would be to monitor progress and contributions made by the program that involve these four processes.

At the same time, the assessment procedures might at first blush appear to be daunting. An evaluation team might have to convene a variety of expert panels, each serving as peer reviewers by covering different disciplines or education topics, and even combinations of both. The presumed panels would review ideas and practices emerging from the MSP Program, similar to the role of judges in other major competitions that culminate in the awarding of prizes (Clinedinst, 2005). In other words, the procedure would emulate that followed in all other academic fields, which use peer review processes to recognize or rebuff claims of discovery or innovation.

TABLE 1  
Four Discovery and Innovation Processes

Process	Preceded by	Illustrative Item (and Item No.) From the Appendix	Rivalled by
1. <i>Uncovering</i>	<ul style="list-style-type: none"> <li>■ Making hunches →</li> <li>■ Searching</li> </ul>	6. Expectation (Pygmalion) Effect	Not new, not important, artifact
2. <i>Inventing</i>	<ul style="list-style-type: none"> <li>■ Diagnosing →</li> <li>■ Collecting/assembling →</li> <li>● Tinkering</li> </ul>	1. Intelligence Testing	Not useful, not better, not novel ("obvious")
3. <i>Explaining</i>	<ul style="list-style-type: none"> <li>■ Conceptualizing →</li> <li>■ Criticizing extant theories →</li> <li>■ Predicting</li> </ul>	9. Time on Task	Not sound, not insightful, not logical, not plausible, not valid, not appropriate (e.g., reductionist)
4. <i>Substantiating</i>	<ul style="list-style-type: none"> <li>■ Testing hypotheses →</li> <li>■ Replicating →</li> <li>■ Conducting meta-analyses</li> </ul>	21. Summer Learning Loss	Spurious, uncontrolled, unsubstantiated, not statistically significant

A more pragmatic perspective suggests an alternative that might work equally well. The needed peer review function is already being performed when researchers make presentations of their work, submit manuscripts for formal publication, or even develop proposals for new funding. The pursuit of discoveries and innovations normally must follow this gradual communication process whereby science is publicly shared (Polanyi, 1967/1983), usually in increasingly wider circles. Moreover, a researcher's presentation and manuscript submissions are only the first (but crucial) step in the review process. Ideas and arguments are sharpened and shaped on the anvil of peer review—softened by the heat of criticism and hammered into form by advice, revision, and reworking. More prosaically, during this process authors are challenged to exclude alternative explanations, run additional experimental controls, and differentiate or integrate arguments and inferences in the literature. Sometimes specific implications are drawn by extension or deduction, and the author is challenged to eliminate them or account for them.

As a result, by the time manuscripts are finally published, the work has been subjected to an initial round of peer review. Subsequent citation, replication,

challenge to the published work, and newly published data and insights constitute a continuing cycle of such review (updating and refinement).

Evaluations, therefore, need not assemble their own review processes but can take advantage of the routine ones already in place. Monitoring formal presentations and publications emanating from an R&D effort can serve as the needed evaluation procedure. For instance, the MSP Program has convened two annual meetings called “evaluation summits,” where papers were formally presented. Other presentations have been made at larger professional gatherings, such as those of the American Educational Research Association. By involving disciplinary and other institution of higher education faculty as core partners, the MSP Program can expect a stream of scholarship published in education-related academic and practitioner journals.

Monitoring these sources for evidence of the four types of discovery and innovation processes—uncovering, inventing, explaining, and substantiating—would be one possible way of assessing the R&D functions of the MSP Program.

## SUMMARY

The MSP Program, consisting of a portfolio of funded projects, in part positions itself as an R&D program. This study has addressed the need to assess how well the R&D function is working, beyond the program’s possible impact on existing pre-K-12 systems.

The study discusses and enumerates discoveries and innovations in education and other fields. In so doing, the study suggests four types of discovery and innovation that can be monitored as part of an assessment of the MSP Program: uncovering, inventing, explaining, and substantiating. The study concludes that the needed R&D assessment can occur by monitoring the funded projects for their formal presentations and publications for evidence of these four types of discovery and innovation.

By focusing on formally presented or published works, such an approach represents what Corley (2006) called a “state-of-the-art” approach. However, Corley believes that the approach is limited because it mainly emphasizes knowledge production, giving inadequate emphasis to the actual application of the ideas and innovations. For the MSP Program, such a limitation may be somewhat attenuated by the fact that the relevant presentations and publications include those of teachers and other educational practitioners, reporting about their implementation experiences. In this sense, and given the realistic confines of a time-limited evaluation, the planned assessment therefore can reflect the fuller breadth of the MSP Program’s potential contributions, covering both knowledge production and knowledge use.

## ACKNOWLEDGMENT

This study is one in a series of substudies for the Math and Science Partnership Program Evaluation (MSP-PE) conducted for the National Science Foundation's MSP. The MSP-PE is conducted under Contract No. EHR-0456995. Since 2007, Bernice Anderson, Ed.D., Senior Advisor for Evaluation, Directorate for Education and Human Resources, has served as the National Science Foundation Program Officer. The authors are Robert K. Yin, of COSMOS Corporation, Edward J. Hackett, of Arizona State University, and Daryl E. Chubin, of American Association for the Advancement of Science.

The MSP-PE is led by COSMOS Corporation in current partnership with George Mason University (GMU) and Brown University. Robert K. Yin (COSMOS) serves as Principal Investigator and Jennifer Scherer (COSMOS) serves as one of three Co-Principal Investigators. Additional Co-Principal Investigators and their collaborating institutions (including discipline departments and math centers) are Patricia Moyer-Packenham (USU, formerly GMU) and Kenneth Wong (Brown).

## REFERENCES

- Bell, D. (1980). *The winding passage*. Cambridge, MA: Abt Books.
- Berliner, D. (2002). Educational research: The hardest science of all. *Educational Researcher*, 31, 18–20.
- Brannigan, A. (1981). *The social basis of scientific discoveries*. Cambridge: Cambridge University Press.
- Clinedinst, M. (2005, September). *Core MSP-PE evaluation design papers: Background information about discovery criteria in other fields (First annual report)* [unpublished paper] COSMOS Corporation, Bethesda, MD.
- Collins, H. M. (1985). *Changing order*. Beverly Hills, CA: Sage.
- Corley, E. A. (2006). A use-and-transformation model for evaluating public R&D: Illustrations from Polycystic Ovarian Syndrome (PCOS) research. *Evaluation and Program Planning*, 30, 21–35.
- Feuer, M. (2004, November 18). *The science of rationality and the rationality of science: Thoughts on the future of education research and policy* (Burton-Inglis Lectures). Cambridge, MA: Harvard Graduate School of Education.
- Fleck, L. (1981). *Genesis and development of a scientific fact*. Chicago: University of Chicago Press.
- Jadad, A. (1998). *Randomized controlled trials*. London: BMJ.
- Kelly, A. E., & Yin, R. K. (2007). Strengthening structured abstracts for educational research: The need for claim-based structured abstracts. *Educational Researcher*, 36(3), 133–138.
- Klein, J. T. (1996). *Crossing boundaries: Knowledge, disciplinarity, interdisciplinarity*. Charlottesville: University Press of Virginia.
- Latour, B., & Woolgar, S. (1979). *Laboratory life: The social construction of scientific facts*. Princeton, NJ: Princeton University Press.
- Merton, R. K. (1973). Singletons and multiples in scientific discovery: A chapter in the sociology of science. In R. K. Merton (Ed.), *The sociology of science: Theoretical and empirical investigation* (pp. 343–370). Chicago: University of Chicago Press.



- National Science Foundation. (2001). *Math and Science Partnership (MSP) Program* (Program solicitation, NSF-02-061). Arlington, VA: Author.
- National Science Foundation. (2003). *Math and Science Partnership Program (MSP) Program* (Program solicitation, NSF-03-605). Arlington, VA: Author.
- Polanyi, M. (1983). *The tacit dimension*. New York: Doubleday. (Original work published 1967).
- Shavelson, R. J., & L. Towne. (Eds.). (2002). *Scientific research in education*. Washington, DC: National Research Council, National Academy Press.
- Stent, G. (2002). Prematurity in scientific discovery. In E. B. Hook (Ed.), *Prematurity in scientific discovery* (pp. 22–36). Berkeley: University of California Press.
- Stokes, D. (1997). *Pasteur’s quadrant: Basic science and technological innovation*. Washington, DC: The Brookings Institution.
- Thagard, P. (1999). *How scientists explain disease*. Princeton, NJ: Princeton University Press.
- Towne, L., & Hilton, M. (Eds.). (2004). *Implementing randomized field trials in education: Report of a workshop*. Washington, DC: Committee on Research in Education, National Research Council.
- Towne, L., Wise, L. L., & Winters, T. M. (Eds.). (2004). *Advancing scientific research in education*. Washington, DC: Committee on Research in Education, National Research Council.
- van Raan, A. F. J. (2004). Sleeping beauties in science. *Scientometrics*, 59(3), 461–466.

## APPENDIX

### Candidate Innovations in Education

#### 1. Intelligence Testing

In 1905, Alfred Binet and Theodore Simon devised a system for testing intelligence, with scoring based on standardized, average *mental levels* for various age groups. In 1916 the Binet-Simon Intelligence Scale was expanded and reworked by Lewis Terman at Stanford University, and later revisions called the Revised Stanford-Binet Intelligence Tests were published in 1937, 1960, and 1985. A highly successful series of tests, designed by psychologist David Wechsler, have been in wide use for years as diagnostic and evaluative instruments. Known in 1939 as the Wechsler-Bellevue Intelligence Scale, the Wechsler Adult Intelligence Scale is a standard tool for intelligence testing today by psychometricians. While no consensus of opinion prevails about what such tests actually measure, their use in education has had great practical value in assigning children to class groups and in predicting academic performance.

For bibliography, see <http://www.infoplease.com/ce6/sci/A0867242.html>

#### 2. The Turing Test

Alan Turing’s 1950 article in *Mind*, “Computing Machinery and Intelligence,” has become one of the most cited in philosophical literature on the development of “artificial intelligence” and computer science education. He asserted verbal ability as a criterion for intelligence: can machine be distinguished from person?

Shieber, S. (Ed.). (2003). *The Turing Test: Verbal behavior as the hallmark of intelligence*. Cambridge, MA: MIT Press. See <http://www.turing.org.uk/turing/scrapbook/test.html> and <http://www.turing.org.uk/turing/>

### 3. Direct Instruction

K-8 math curriculum developed in 1960s: teacher script, student skills mastered and combined in sequence (called DISTAR for K-3 and Connecting Math Concepts for 1-6).

Recognized by BEST (Building Engineering and Science Talent). (2004). *What it takes: Pre-K-12 design principles to broaden participation in science, technology, engineering and mathematics*. Appendix 2, 41-42. Also see The Urban Institute (2005, February). *What do we know? Seeking effective math and science instruction* (pp. 8-9); and Siegfried Engelmann, <http://www.naschools.org/uploadedfiles/CSSP%27s%20award%20for%20Engelmann.pdf>

### 4. Project Seed

Begun in 1963 as Grade 3 to 6 supplement, Special Elementary Education for the Disadvantaged in urban school districts; focused on algebraic concepts, critical thinking, and problem-solving skills; combines teacher PD with parent workshops.

Noted in BEST, *What It Takes*, pp. 43-44. Also see <http://www.projectseed.org/> and <http://www.chemistry.org/portal/a/c/s/1/acsdisplay.html?DOC=education%5Cstudent%5Cprojectseed.html>

### 5. Head Start

In 1965, the Office of Economic Opportunity launched Project Head Start as an eight-week summer program. Head Start was part of the War on Poverty, designed to help break the "cycle of poverty" by providing preschool children of low-income families with a comprehensive program to meet their emotional, social, health, nutritional, and psychological needs. Head Start has grown to include full day/year services and many program options. In the mid-1990s, birth to age 3 services were formalized and expanded with the inception of Early Head Start. A bibliography of more than 3,000 studies documents the impact of Head Start.

<http://www.acf.hhs.gov/programs/hsb/publications/index.htm>

### 6. Expectation Effect (Pygmalion Effect)

During the 1964-65 school year, Robert Rosenthal conducted an experiment in an elementary school to see whether teacher expectations influenced their

students’ performances. Teachers’ expectations indeed improved the academic performance of their students. Additional experiments with laboratory mice labeled as “maze-bright” actually performed better. The “self-fulfilling prophecy” lives in experimenter outcome bias.

Rosenthal, R. (1991). Teacher expectancy effects: A brief update 25 years after the Pygmalion experiment. *Journal of Research in Education*, 1, 3–12. Also see [http://www.psichi.org/pubs/articles/article\\_121.asp](http://www.psichi.org/pubs/articles/article_121.asp)

### 7. Math anxiety

Not a failure of intellect, but a failure of nerve (especially in girls) that can be overcome.

Tobias, S. (1994). *Overcoming math anxiety*. New York: Norton. (Original work published 1978). Also see Math Anxiety: Internet Resources, [www.oncourseworkshop.com/Emotions006.htm](http://www.oncourseworkshop.com/Emotions006.htm)

### 8. Treisman’s Collaborative Learning Model

A math workshop model of the late 1970s pioneered at Cal-Berkeley to combat the high failure rates of minority students in undergraduate calculus courses. Build a community around the study of mathematics in which problem sets drove the group interaction and students’ strengths are emphasized.

Treisman, U.(1992). Studying students studying calculus: A look at the lives of minority mathematics students in college. *The College Mathematics Journal*, 23, 362–372. Also see [www.math.uiuc.edu/MeritWorkshop/uriModel.html](http://www.math.uiuc.edu/MeritWorkshop/uriModel.html)

### 9. Time on Task

In the mid-1970s, investigations of effective teaching led Jane Stallings and her colleagues to develop the classroom snapshot, a quantitative instrument that records how time is spent in class. Her research led to a shift to a more effective use of time or active teaching and learning. Although subsequent research found Stallings recommendation beneficial for reading instruction, the research also suggested adapting different techniques for instruction to other subjects, such as mathematics.

Stallings, J. (1980). Allocated academic learning time revisited, or beyond time on task. *Educational Leadership*, 9(11), 11–16.

### 10. The Algebra Project

Local community and national networks begun in 1982 to assist students of color to complete algebra by 9th grade and calculus by 12th.

Moses, R. P. (1994). Remarks on the struggle for citizenship and math/sciences literacy. *Journal of Mathematical Behavior*, 13, 107–111.

Also see [www.algebra.org/apinfo/over2.html](http://www.algebra.org/apinfo/over2.html)

### 11. Cultural Capital

French sociologist Pierre Bourdieu asserted that the maintenance of a system of power by means of the transmission of a dominant culture. One of the central themes in his works was that culture and education affirm differences between social classes and in the reproduction of those differences. Randall Collins's later work on the "credential society" is an elaboration—credentials are important in themselves and not for the skills their bearer has acquired.

See <http://www.kirjasto.sci.fi/bourd.htm>

### 12. Computer Assisted Instruction

In the 1980s, computer use in classrooms followed multiple studies testing and demonstrating potential student benefits. The research also allayed fears that teachers' jobs would not be jeopardized. The more substantive lines of research showed how classroom computers—initially employed for fairly rigid drill and practice—could provide tutorial support more responsive to individual student needs.

Dalton, D. W., & Hannafin, M. J. (1988). The effects of computer-assisted and traditional mastery methods on computation accuracy and attitudes. *Journal of Educational Research*, 82(1), 27–33.

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. C. (1985). Effectiveness of computer-based education in elementary schools. *Computers in Human Behavior*, 1(1), 59–74.

### 13. Constructivism

In the 1960s, researchers sought to understand mathematical reasoning in children. Their investigations, grounded in the work of Jean Piaget, led to a realization that students are active learners who construct a set of conceptual structures that constitute a personal knowledge base. The role of the teacher then is not a "sage on the stage, but a guide on the side." The National Science Foundation's Directorate for Education and Human Resources, among others, supported multiple studies that eventually showed how "constructivism" could be applied in the K-12 classroom. Today, constructivism undergirds current understanding of higher order thinking and problem-solving skills, authentic work, active learning, student-directed and student-guided instruction, and deep work.

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher*, 23(7), 13–20.



COSMOS Corporation. (1999). *Advancing constructivism in mathematics education research* (Final Report on NSF’s Research on Education, Policy, and Practice Program). Bethesda, MD: Author.

#### *14. Interdisciplinary Science, Mathematics, Engineering, and Technology (STEM) Education*

The 1990s ushered in an era of hybridization where disciplinary boundaries blurred because of research advances. NSF supported much of this “interdisciplinarity” through the funding of centers. As but one of these initiatives, early in the 21st-century, “Science of Learning Centers” (SLCs) were established as multi-institutional, multiyear consortia often destined for absorption into the campus structure. Notably, the ultimate focus of SLCs is learning and teaching. For example, CELEST—A Center for Learning in Education, Science, and Technology funded for \$9M across 5 years—brings together researchers from Boston University, Brandeis, MIT, and the University of Pennsylvania to study real-time autonomous learning systems by integrating experimental and computational brain science, biologically inspired technology, and classroom innovation. Activities center on

- Quantitative behavioral and brain modeling of normal and abnormal learning
- Interdisciplinary cognitive and neuroscience experiments to probe these processes and test hypotheses
- Development of algorithms, based on biological learning models, for fast learning about complex and rapidly changing environments in large-scale science and engineering applications
- Integration of research with education through contributions to educational technology and curriculum development

Once degree programs emerge in fledgling fields, then institutionalization is more of a reality. The scholarship that underlies this process can be seen at <http://cns.bu.edu/celest/> Through these centers, NSF seeks “to create the intellectual, organizational, and physical infrastructure needed for the long-term advancement of learning research.”

[http://www.nsf.gov/funding/pgm\\_summ.jsp?pims\\_id=5567&from=fund](http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5567&from=fund)

#### *15. Value-Added Assessment*

Implemented in 1992 using a complex statistical method of measuring student growth via tests in Grades 2 to 8.

Sanders, W. L., & Horn, S. (1994). The Tennessee Value-Added Assessment System (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, 8, 299–311. For a bibliography,

see [http://www.schoolwisepress.com/smart/browse/browse\\_val.html](http://www.schoolwisepress.com/smart/browse/browse_val.html). For a related higher education approach linking assessment to accountability spearheaded by RAND/Council for Aid to Education, see Benjamin and Hersh at <http://www.aacu-edu.org/peerreview/pr-sp02/pr-sp02feature2.cfm>

### 16. Multiple Intelligences

In the introduction to the 10th anniversary edition of his classic, *Frames of Mind. The Theory of Multiple Intelligences*, Howard Gardner wrote

In the heyday of the psychometric and behaviorist eras, it was generally believed that intelligence was a single entity that was inherited; and that human beings—initially a blank slate—could be trained to learn anything. . . . Nowadays an increasing number of researchers believe precisely the opposite; that there exists a multitude of intelligences, quite independent of each other; that each intelligence has its own strengths and constraints; that the mind is far from unencumbered at birth; and that it is unexpectedly difficult to teach things that go against early “naïve” theories or that challenge the natural lines of force within an intelligence and its matching domains. (p. xxiii)

Gardner’s work has been called a “paradigm shifter,” challenging the notion of a single IQ test and the cognitive development work of Piaget.

<http://www.infed.org/thinkers/gardner.htm>

### 17. Stereotype Threat

The threat of being viewed through the lens of a negative stereotype or the fear of doing something that would inadvertently confirm that stereotype, demonstrated in 1995 with African American students taking a standardized test; since confirmed with all groups.

Steele, C. M., & Aronson J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797–811. Also see [www.personal.psu.edu/users/t/r/trc139/references.htm](http://www.personal.psu.edu/users/t/r/trc139/references.htm)

### 18. World Wide Web

The World Wide Web has myriad applications in K-12 and higher education, including teacher professional development, curriculum materials, and Internet-based information on particular concepts and tools. Emerging in the last 15 years as “distance learning,” these resources can be accessed online under such terms as global learning communities, *Science NetLinks*, and virtual dissection. Thousands of education resources are indexed at <http://www.educationindex.com/educator/>

For advancing higher education through the use of information technologies, see [http://www.educause.edu/content.asp?PAGE\\_ID=720&bhcp=1](http://www.educause.edu/content.asp?PAGE_ID=720&bhcp=1)

### 19. K-16 Systemic Education Reform

The Puerto Rico Statewide Systemic Initiative has enhanced student performance in mathematics, improving access to higher education institutions, and motivating more to pursue STEM careers. It has demonstrated how a more seamless K-16 system, coupling education levels through various systemic initiatives and unifying funding streams, can support the achievement of all students. The chief architect, physicist Manuel Gomez, applied systems theory to redesigning math and science education as vice president of the University of Puerto Rico at Rio Piedras. [www.crci.uprr.pr/rcse/About%20CSE.htm](http://www.crci.uprr.pr/rcse/About%20CSE.htm)

### 20. Success for All

A national randomized field trial launched in 2001 finds that K-1 students read better after 2 years in the program. Success for All was pioneered in reading and mathematics by Robert Slavin.

Slavin, R. E., & Madden, N. A. (1999). *Success for All/Roots and Wings: 1999 summary of research on achievement outcomes*. <http://www.successforall.net/resource/research/report41entire.pdf>

Also see [www.edweek.org/ew/articles/2005/05/11/36success.h24.html?rale=14RcsgF70mPtC](http://www.edweek.org/ew/articles/2005/05/11/36success.h24.html?rale=14RcsgF70mPtC) and <http://www.ecs.org/clearinghouse/18/93/1893.htm>

### 21. Summer Learning Loss (and Regression)

The phenomenon of how student learning regresses over summer vacation, requiring time at the start of the following school year to recover from the loss.

Alexander, P., Entwistle, D. R., & Olson, L. S. (2004). Schools, achievement, and inequality: A seasonal perspective. In G. Borman & M. Boulay (Eds.), *Summer learning: Research, policies, and programs*. Mahwah, NJ: Erlbaum.

Also see <http://www.ericdigests.org/2003-5/summer.htm>

### 22. Bloom's Taxonomy

In 1948, educators, led by Benjamin Bloom, developed a classification system for three learning domains: cognitive, psychomotor, and affective. In 1956, the cognitive domain was defined for six levels ranging from basic to more complex (i.e., knowledge, comprehension, application, analysis, synthesis, and evaluation). Research has confirmed the hierarchy with the exception of the last two levels with

some arguing that they are equally complex, whereas others suggest they should be reversed. In any case, the taxonomy is widely used in developing test items and in assessing the rigor of education goals and objectives.

Bloom, B., Englehart, M., Furst, E., Hill, W., & Krathwohl, D. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York: Longmans.

Huitt, W. (2004). Bloom et al.'s taxonomy of the cognitive domain. *Educational Psychology Interactive*. Valdosta, GA: Valdosta State University.



ISSN: 0161-956X

# Peabody Journal of Education

Volume 83  
2008



**Routledge**  
Taylor & Francis Group

# PEABODY JOURNAL OF EDUCATION

Camilla P. Benbow, Dean, Peabody College  
James W. Guthrie, Editor  
Michele Thompson, Coordinating Editor  
Deborah M. Enright, Associate Editor

## EDITORIAL ADVISORY BOARD

Darlyne Bailey, *University of Minnesota*  
R. Dale Ballou, *Peabody College of Vanderbilt University*  
Sir Clive Booth, *Oxford Brookes University, England*  
Mary M. Brabeck, *New York University*  
Mark Bray, *The University of Hong Kong, PRC*  
Louis A. Castenell, Jr., *The University of Georgia*  
David Y. Chen, *Chongquin Technology and Business University (CTBU)*  
Mike Dishman, *McLocklin, Murphy & Dishman, L.L.P.*  
Lora Cohen-Vogel, *Florida State University*  
Joseph P. Du Cette, *Temple University*  
Victor Firsov, *Russian Academy of Teachers Retraining, Russia*  
Juan Froemel, *LLECE/UNESCO, Chile*  
Louis M. Gomez, *Northwestern University*  
Gerardo M. Gonzalez, *Indiana University*

David W. Gordon, *Elk Grove Unified School District, California*  
Eric A. Hanushek, *Hoover Institution, Stanford University*  
Stephen P. Heyneman, *Peabody College of Vanderbilt University*  
Phillip W. Jones, *The University of Sydney, Australia*  
Kazuo Kuroda, *Waseda University, Japan*  
James H. Lytle, *Trenton Public Schools, New Jersey*  
Michael K. McLendon, *Peabody College of Vanderbilt University*  
Ann L. Marcus, *New York University*  
Bill McKinney, *Region IV Education Service Center, Texas*  
Francois Orivel, *University of Bourgogne, France*  
Penelope L. Peterson, *Northwestern University*  
Michael J. Podgursky, *University of Missouri-Columbia*  
Robert Slavin, *Johns Hopkins University*  
Jandhyala B. G. Tilak, *National Institute of Education Planning and Administration, India*

---

**Abstracted and/or Indexed in:** PsycINFO/Psychological Abstracts; Linguistics and Language Behavior Abstracts; Contents Pages in Education; Sociological Abstracts; Education Index; Education Abstracts; Social Services Abstracts; EBSCOhost Products; Cabell's Directories; Family Index Database.

*Peabody Journal of Education* (ISSN: 0161-956X) is published quarterly in February, May, August, and November for a total of 4 issues per year by Taylor and Francis Group, LLC, 325 Chestnut Street, Suite 800, Philadelphia, PA 19106.

**US Postmaster:** Please send address changes to Peabody Journal of Education, Taylor & Francis Group, LLC, 325 Chestnut Street, Suite 800, Philadelphia, PA 19106.

### Annual Subscription, Volume 83, 2008

Print ISSN - 0161-956X, Online ISSN - 1532-7930

Institutional subscribers: US \$462, UK £277, €370

Personal subscribers: US \$58, UK £35, €46

Institutional and individual subscriptions include access to the online version of the journal. Institutional subscriptions include access for any number of concurrent users across a local area network. Individual subscriptions are single username/password only.

**Production and Advertising Office:** 325 Chestnut Street, Suite 800, Philadelphia, PA 19106. Tel: 215-625-8900, Fax: 215-625-8563. Production Editor: Michelle McCarthy

### Subscription offices

**USA/North America:** Taylor and Francis Group, LLC, 325 Chestnut Street, Suite 800, Philadelphia, PA 19106. Tel: 215-625-8900, Fax: 215-625-2940.

**UK/Europe:** Taylor and Francis Customer Services, Sheepen Place, Colchester, Essex CO3 3LP, UK. Tel: +44 (0)-20-7017-5544, Fax: +44 (0)-20-7017-5198.

For a complete guide to Taylor and Francis Group, LLC's journal and book publishing programs, visit our website: [www.taylorandfrancis.com](http://www.taylorandfrancis.com)

**Copyright** © 2008 Taylor and Francis Group, LLC. All rights reserved. No part of this publication may be reproduced, stored, transmitted, or disseminated in any form or by any means without prior written permission from Taylor and Francis Group, LLC. Taylor and Francis Group, LLC grants authorization for individuals to photocopy copyright material for private research use. The publisher assumes no responsibility for any statements of fact or opinion expressed in the published papers. The appearance of advertising in this journal does not constitute an endorsement or approval by the publisher, the editor, or the editorial board of the quality or value of the product advertised or of the claims made for it by its manufacturer.

**Permissions.** For further information, please visit: <http://www.tandf.co.uk/journals/permissions.asp>

November 2008

# Peabody Journal of Education

## Table of Contents for Volume 83

Volume 83, Number 1, 2008

Introduction to the Special International Issue <i>Stephen P. Heyneman</i>	1
The Origins of the International Standard Classification of Education <i>John A. Smyth</i>	5
Implementing Free Primary Education Policy in Malawi and Ghana: Equity and Efficiency Analysis <i>Kazuma Inoue and Moses Oketch</i>	41
Globalization and Implementation of an Equity Norm in Higher Education: Admission Processes and Funding Framework Under Scrutiny <i>Gaële Goastellec</i>	71
Aspects of Fiscal Federalism in Higher Education Cost Sharing in Latvia <i>Rita Kasa</i>	86
Higher Educational Cost-Sharing, Dual-Track Tuition Fees, and Higher Educational Access: The East African Experience <i>Pamela Marcucci, D. Bruce Johnstone, and Mary Ngolovoi</i>	101
Do Electronic Technologies Increase or Narrow Differences in Higher Education Quality Between Low- and High-Income Countries? <i>Norman Clark Capshaw</i>	117
Compulsion, Craft, or Commodity? Education Services Trade in the Larger Context <i>Brandyn L. Payne</i>	133
Diagnosis, Treatment, and Educational Implications for Students With Attention-Deficit/Hyperactivity Disorder in the United States, Australia, and the United Kingdom <i>Sarah Schlachter</i>	154

## Volume 83, Number 2, 2008

Introduction to the Special Issue on From Equity to Adequacy to Choice <i>Michael Podgursky</i>	171
A New Defendant at the Table: An Overview of Missouri School Finance and Recent Litigation <i>Michael Podgursky, James Smith, and Matthew G. Springer</i>	174
What Do Cost Functions Tell Us About the Cost of an Adequate Education? <i>Robert Costrell, Eric Hanushek, and Susanna Loeb</i>	198
Can Judges Improve Academic Achievement? <i>Jay P. Greene and Julie R. Trivitt</i>	224
Spending Money When It Is Not Clear What Works <i>Paul T. Hill</i>	238
Next Needed Steps in the Evolution of American Education Finance and Policy: Attenuating a Judicially Imposed Policy Distraction, Activating a Balanced Portfolio of K-12 School Reforms, Advancing Rationality as a Goal in Pursuing Productivity, Advocating Change in a Responsible and Effective Manner <i>James W. Guthrie</i>	259
A Framework for Choice Remedy Litigation <i>Clint Bolick</i>	285
Hope for Children Trapped in Failing Schools: The Promise of <i>Crawford v. Davy</i> <i>Julio C. Gomez, Esq.</i>	297

## Volume 83, Number 3, 2008

Introduction to the Special Issue on Scaling Up Teaching and Learning Improvement in Urban Districts: The Promises and Pitfalls of External Assistance Providers <i>Amanda Datnow and Meredith I. Honig</i>	323
Adaptive Assistance for Learning Improvement Efforts: The Case of the Institute for Learning <i>Meredith I. Honig and Gina S. Ikemoto</i>	328
Authority, Status, and the Dynamics of Insider-Outsider Partnerships at the District Level <i>Cynthia E. Coburn, Soung Bae, and Erica O. Turner</i>	364



Collaborative Assistance in a Highly Prescribed School Reform Model: The Case of Success for All <i>Vicki Park and Amanda Datnow</i>	400
Assistance and Accountability in Externally Managed Schools: The Case of Edison Schools, Inc. <i>Julie Marsh, Laura Hamilton and Brian Gill</i>	423
Melding Internal and External Support for School Improvement: How the District Role Changes When Working Closely With External Instructional Support Providers <i>Jonathan Supovitz</i>	459

### Volume 83, Number 4, 2008

Introduction to the MSP-PE Special Issue on Math and Science Partnership Program: A First Comprehensive Evaluation <i>Kenneth K. Wong, Robert K. Yin, Patricia S. Moyer-Packenham, and Jennifer Scherer</i>	479
The Math and Science Partnership Program Evaluation: Overview of the First Two Years <i>Robert K. Yin</i>	486
A Review of the Literature on Mathematics and Science Teacher Quality <i>Johnna J. Bolyard and Patricia S. Moyer-Packenham</i>	509
Does MSP Participation Increase the Supply of Math Teachers? Developing and Testing an Analytic Model <i>John H. Tyler and Svetla Vitanova</i>	536
The Assessment of Mathematics and Science Teacher Quality <i>Patricia S. Moyer-Packenham, Johnna J. Bolyard, Anastasia Kitsantas, and Hana Oh</i>	562
Mathematics Curriculum Systems: Models for Analysis of Curricular Innovation and Development <i>Margret A. Hjalmarson</i>	592
A Review of Instruments to Evaluate Partnerships in Math and Science Education <i>Jennifer Scherer</i>	611
Initial Trends in MSP-Related Changes in Student Achievement With MIS Data <i>Dimitar M. Dimitrov</i>	637

# INSTRUCTIONS FOR AUTHORS

**Content:** *Peabody Journal of Education (PJE)* publishes quarterly symposia in the broad area of education, including but not limited to topics related to formal institutions serving students in early childhood, preschool, primary, elementary, intermediate, secondary, postsecondary, and tertiary education. The scope of the journal includes special kinds of educational institutions, such as those providing vocational training or the schooling for students with disabilities. PJE also welcomes immediate framework of institutions, and education matters that are important to nations outside the United States. Finally, it includes topics that are linked to the social and organizational context in which formal and informal education take place. The Editor cooperates with groups of scholars to present multifaceted, integrated expositions of important topics. A given issue of PJE may contain contributions from social scientists, historians, philosophers, attorneys, practitioners, and policymakers. Unsolicited proposals for special issues - including designation of participating scholars and an outline of articles - will be accepted for review. Additionally, the Editor cooperates with Editorial Board members to identify potential topics, guest editors, and contributors. PJE has the flexibility to consider publishing monographs or a series focused on particular lines of inquiry. In all cases, the Editor and the Editorial Board will ensure that each issue is carefully reviewed and its articles will comprise a high-quality contribution to understanding and practice.

**Manuscript Preparation:** To submit a prospectus for a special issue, note submission requirements presented here and contact the Coordinating Editor, Michele Thompson, for complete details (address below). Prepare manuscripts according to the *Publication Manual of the American Psychological Association* (5th ed., 2001; available from APA Order Department, P.O. Box 2710, Hyattsville, MD 20874), especially with regard to reference lists and text citations. Follow Guidelines to Reduce Bias in Language (*APA Manual*, pp. 61–87). Manuscripts should include the theme title and rationale for the theme. Using 8.5 × 11-in., nonsmear paper and 1-in. margins, type all components double spaced and in the following order: title page (p. 1), abstract (p. 2), text (including quotations), references, appendices, footnotes, tables, and figure captions. On page 1, type article title, author name(s) and affiliation(s), running head (abbreviated title, no more than 45 characters and spaces), author notes and acknowledgments, and name and address of person to whom requests for reprints should be addressed; on page 2, type an abstract of no more than 150 to 200 words. Indent all paragraphs. Use footnotes sparingly. Attach photocopies of all figures. Number all manuscript pages (including photocopies of figures).

**Cover Letter, Permissions, Credit Lines:** In a cover letter, include the contact author's complete mailing address, e-mail address, and telephone and fax numbers. State that the manuscript includes only original material that has not been published and that is not under review for publication elsewhere. Authors are responsible for all statements made in their work and for obtaining permission to reprint or adapt a copyrighted table or figure or to quote at length from a copyrighted work. Authors should write to original author(s) and original publisher to see if permission is required and to request nonexclusive world rights in all languages to use the material in the current article and in future editions. Include copies of all permissions and credit lines with the manuscript. (See p. 175 of the *APA Manual* for samples of credit lines.)

**Manuscript and Computer Disk Submissions:** Prepare manuscript on a word processor and submit three (3) high-quality printouts and a disk to the Coordinating Editor, Michele Thompson, *Peabody Journal of Education*, Department of Leadership, Policy, and Organizations, Peabody College of Vanderbilt University, Peabody #41, 230 Appleton Place, Nashville, TN 32703-5721. The disk should contain two files - word processor and ASCII versions of the manuscript - and file content must match the printed manuscript exactly, or there will be a delay in publication. Manuscripts and disks are not returned.

**Accepted Manuscripts:** Authors are strongly encouraged to submit manuscripts on disk. The disk should be prepared using MS Word or WordPerfect and should be clearly labeled with the authors' names, file name, and software program. A hardcopy printout that exactly matches the disk must be supplied. Each manuscript must be accompanied by a statement that it has not been published elsewhere and that it has not been submitted simultaneously for publication elsewhere. Authors are responsible for obtaining permission to reproduce copyrighted material from other sources and are required to sign an agreement for the transfer of copyright to the publisher. All accepted manuscripts, artwork, and photographs become the property of the publisher. All parts of the manuscript should be typewritten, double-spaced, with margins of at least one inch on all sides. Number manuscript pages consecutively throughout the paper. Authors should also supply a shortened version of the title suitable for the running head, not exceeding 50 character spaces. Each article should be summarized in an abstract of not more than 100 words. Avoid abbreviations, diagrams, and reference to the text in the abstract.

**Illustrations:** Illustrations submitted (line drawings, halftones, photos, photomicrographs, etc.) should be clean originals or digital files. Digital files are recommended for highest quality reproduction and should follow these guidelines:

- 300 dpi or higher
- Sized to fit on journal page
- EPS, TIFF, or PSD format only
- Submitted as separate files, not embedded in text files

**Color Reproduction.** Color illustrations will be considered for publication; however, the author will be required to bear the full cost involved in color art reproduction. Color art can be purchased for online only conversion and reproduction or for print + online reproduction. Color reprints can only be ordered if print + online reproduction costs are paid. Rates for color art reproduction are: **Online Only Reproduction:** \$225 for the first page of color; \$100 per page for each of the next three pages of color. A maximum charge of \$525 applies. **Print + Online Reproduction:** \$900 for the first page of color; \$450 per page for the next three pages of color. A custom quote will be provided for articles with more than 4 pages of color.

**Tables and Figures.** Tables and figures (illustrations) should not be embedded in the text, but should be included as separate sheets or files. A short descriptive title should appear above each table with a clear legend and any footnotes suitably identified below. All units must be included. Figures should be completely labeled, taking into account necessary size reduction. Captions should be typed, double-spaced, on a separate sheet. All original figures should be clearly marked in pencil on the reverse side with the number, author's name, and top edge indicated.

**Proofs and Reprints.** Page proofs are sent to the designated author using Taylor & Francis' Central Article Tracking System (CATS). They must be carefully checked and returned within 48 hours of receipt. Reprints of individual articles are available for order at the time authors review page proofs. A discount on reprints is available to authors who order before print publication.

(Continued from back cover)

Discovering “What’s Innovative”: The Challenge of Evaluating  
Education Research and Development Efforts

674

*Robert K. Yin, Edward J. Hackett and Daryl E. Chubin*

**Following Page 690**

Title page for Volume 83

Table of Contents for Volume 83

Author Index for Volume 83







